Photoresponsive ligands for G-quadruplex DNA

Michael P. O'Hagan

A dissertation submitted to the University of Bristol in accordance with the requirements for the award of the degree of Doctor of Philosophy in the Faculty of Science

School of Chemistry, May 2020

Supervisors:

Professor M. Carmen Galan (School of Chemistry, University of Bristol, UK)

Dr Juan Carlos Morales (Institute of Parasitology and Biomedicine, CSIC, Spain)

53,527 words

Abstract

G-quadruplexes are four-stranded nucleic acid secondary structures formed from sequences rich in guanine. These structures display potential as therapeutic targets and as the basis of functional materials. Both of these applications rely on exerting control over the formation of these polymorphic architectures. Small molecule ligands can exert this control by selective stabilisation of the folded structures, but normally only allow a one-way response as their delivery to a system cannot be readily reversed. Photoresponsive ligands offer an opportunity to circumvent this limitation, since their activity may be switched on or off with spatiotemporal precision based on a change in the ligand structure upon irradiation with light. In therapy, this level of control could allow the activity of a G-quadruplex targeting drug to be restricted to a desired site of action, such as a tumour, in an approach known as photopharmacology. Meanwhile, the possibility to control nucleic acid/ligand assemblies reversibly could be deployed in the development of new responsive materials. Despite these opportunities, scant attention to date has been devoted towards the development of photoresponsive G-quadruplex targeting molecules.

This thesis reports the investigation of two photoresponsive scaffolds, stiff-stilbene and dithienylethene, as the basis of novel G-quadruplex ligands (Figure 1). The design, synthesis and optimisation of candidate ligands derived from these chromophores is described. The ligands are evaluated as G-quadruplex binding agents by means of fluorescence-based melting assays, UV/visible spectroscopy, circular dichroism spectroscopy, NMR spectroscopy and molecular dynamics simulations. The bioactivity of the ligands against a small panel of mammalian cells and parasitic organisms is also investigated. The photochemistry of the ligands in physiologically-relevant conditions is explored and the respective photoresponses deployed to control G-quadruplex structure, ligand binding mode and cytotoxicity.

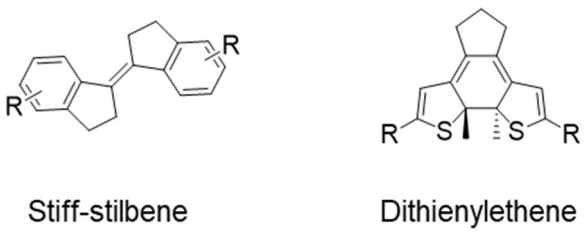


Figure 1: G-quadruplex ligand scaffolds explored in this work.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: …… DATE: ……

Acknowledgements

First of all, my thanks go to my supervisor, Professor M Carmen Galan. Carmen has been a wonderful supervisor throughout my whole PhD, allowing me the freedom to choose my own project and the space to develop it, but always being on hand to offer helpful advice and guidance where I have needed it. She has encouraged me through the tough moments and helped me celebrate the successful ones, and her approach to work is inspirational. Thank you, Carmen, for everything.

I also thank my second supervisor, Dr Juan Carlos Morales, for welcoming me into his lab in Granada for my research stay and for all his scientific guidance, both during my time there and over Skype (before Zoom was a thing!) when back in the UK. My gratitude also goes to Pablo Peñalver for all his help and patience in training me in the biological aspects of the project. Muchas gracias chicos!

Many thanks to all the Galan group past and present for making the lab a happy place to be. Thanks especially to the Dalton's Dollar Savers crew and, in particular, to Sadiyah Sheikh and Robin Jeanneret for proofreading parts of this thesis. I am indebted to Steven Street for starting up nucleic acids research in the group and helping me take my first steps with the project. In later days, it was a pleasure to collaborate with Javier Ramos Soriano to establish further project directions; I am excited to see where the group take these projects next.

I am very grateful to Marta Duchi for help with photochemistry (and homecooked lunches), Susanta Haldar for undertaking computational studies, and Rosie Gibson for her contributions to this project during her MSc studies.

Thank you to the management of the Bristol Chemical Synthesis CDT (Professor Kevin Booker-Milburn, Emma Rose, Mar Ruiz-Molina and Laura Chavda) for the opportunity to undertake my PhD and for their efforts to develop such a successful and engaging training programme. I also thank all the technicians and support staff at the School of Chemistry for their sterling work that allows research at the department to flourish. Tim Harrison runs a great outreach programme and it has been a privilege contribute to some of these activities over the years. Tim's enthusiasm for kitchen chemistry is infectious.

Outside of the lab, I thank my many very good friends who make life interesting and fun, particularly the ones who challenge and develop me the most. I would love to name you all, but the full list would need a whole chapter of its own. Above all else, I express my heartfelt love to my family, my parents Rosie and Bernie and my brother Daniel. You are always there whenever I need you, and this 'lockdown thesis' would not have been possible without you.

Publications

Much of work in this thesis is based on the following publications of which I am first author:

Chapter 1

M.P. O'Hagan, J.C. Morales and M.C. Galan (2019) Binding and beyond: what else can G-quadruplex ligands do? *European Journal of Organic Chemistry*. *31/32,* pp. 4995-5017. [Review article].

Chapter 2 & 3

M.P. O'Hagan, P. Peñalver, R.L.S. Gibson, J.C. Morales and M.C. Galan (2020) Stiff-stilbene ligands target G-quadruplex DNA and exhibit selective anticancer and antiparasitic activity. *Chemistry: A European Journal. 26*, pp. 6224-6233.

Chapter 3 & 4

M.P. O'Hagan, S. Haldar, M. Duchi, T.A.A. Oliver, A.J. Mulholland, J.C. Morales and M.C. Galan (2019) A photoresponsive stiff-stilbene ligand fuels the reversible unfolding of G-quadruplex DNA. *Angewandte Chemie International Edition. 58* (13), pp. 4334-4338.

Chapter 5

M.P. O'Hagan, J. Ramos Soriano, S. Haldar, S. Sheikh, J.C. Morales, A.J. Mulholland and M.C. Galan (2020) Visible-light photoswitching of ligand binding mode suggests G-quadruplex DNA as a target for photopharmacology. *Chemical Communications. 56*, pp. 5186-5189.

I have also contributed to the following additional publications during the course of my PhD studies:

M. Duchi, **M.P. O'Hagan**, R. Kumar, S. Bennie, M.C. Galan, B. Curchod and T.A.A. Oliver (**2019**) Exploring ultraviolet photoinduced charge-transfer dynamics in a model dinucleotide of guanine and thymine. *Physical Chemistry Chemical Physics*. *21*, pp. 14407-14417.

M.P. O'Hagan, J.-L. Mergny and Z.A.E. Waller (**2018**) G-quadruplexes in Prague: a Bohemian Rhapsody. *Biochimie.* 147, pp. 170-180. [Conference report]

Contents

Abbreviations and symbols

Abs	Absorbance
A	Adenine
Å	Ångstrom
Ac	Acetate
Ar	Aryl
ATCC	American Type Culture Collection
AZT	Azidothymidine
c	Concentration
C	Cytosine
CB7	Cucurbit[7]uril
CD	Circular dichroism
cm	Centimetre
COSY	Correlation spectroscopy
DAPI	4′,6-diamidino-2-phenylindole
DCM	Dichloromethane
DFT	Density functional theory
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
dmol	Decimole
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DOTA	1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid
DPPF	1,1'-Bis(diphenylphosphino)ferrocene
DTE	Dithienylethene
E	*Entgegen* (German: opposite)
EC_{50}	Half-maximal effective concentration
EI	Electron ionisation
ELSD	Evaporative light scattering detector
Eqn.	Equation
Equiv.	Equivalents
ESI	Electrospray ionisation
ESR	Electron spin resonance
Et	Ethyl
FAM	Fluorescein
FBS	Fetal bovine serum
FCC	Flash column chromatography
FES	Free energy surface
FRET	Fluorescence resonance energy transfer
G	Guanine
G4	G-quadruplex
h	Hours
HAT	Human African Trypanosomiasis

HEG	Hexaethyleneglycol
HEK	Human embryonic kidney
HMBC	Heteronuclear multiple bond correlation
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum correlation
hv	Irradiation
IC_{50}	Half-maximal inhibitory concentration
ICD	Induced circular dichroism
K	Kelvin
K_a	Association constant
K_B	Boltzmann constant
kcal	Kilocalories
k_{isom}	rate constant of isomerisation
l	Path length (in absorbance spectroscopy)
L. major	*Leishmania major*
LRMS	Low resolution mass spectrometry
M	Molar
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionisation
MD	Molecular dynamics
Me	Methyl
mg	Milligrams
MHz	Megahertz
min	Minutes
mL	Millilitres
mM	Millimolar
mm	Millimetres
mmHg	Millimetres of mercury
mol	Moles
MRC	Medical Research Council
ms	Milliseconds
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
mW	Milliwatts
NAD	Nicotinamide adenine dinucleotide
NDI	Naphthalene diimide
ng	Nanograms
nm	Nanometres
nM	Nanomolar
NMR	Nuclear magnetic resonance
nOe	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
ns	Nanoseconds
PBS	Phosphate-buffered saline
PDB	Protein Data Bank

Ph	Phenyl
ppm	Parts per million
ps	Picoseconds
PSS	Photostationary state
r.t.	Room temperature
R^2	Coefficient of determination
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
RNA	Ribonucleic acid
s	Seconds
sat.	Saturated
SDS	Sodium dodecyl sulfate
T	Temperature
T	Thymine
T. brucei	*Trypanosoma brucei*
TAMRA	5-Carboxytetramethylrhodamine
^{t}Bu	*tert*-Butyl
TEMP	2,2,6,6-Tetramethylpiperidine
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
T_m	Melting temperature
TMPyP4	5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine
TRAP	Telomerase repeated amplification protocol
UV	Ultraviolet (light)
vis	Visible (light)
WTMetaD	Well-tempered metadynamics
Z	*zusammen* (German: together)
° C	Degrees Celsius
δ	Chemical shift
ΔAbs	Change in absorbance
Δδ	Change in chemical shift
ΔT_m	Change in melting temperature
ε	Molar extinction coefficient
θ	Ellipticity
[θ]	Molar ellipticity
λ	Wavelength
µL	Microlitres
µmol	Micromoles
µM	Micromolar
µm	Micrometres

1 Introduction

This chapter is based in part on a review published in the *European Journal of Organic Chemistry*, of which I am co-corresponding author.[1]

This thesis reports progress towards the development of new supramolecular tools for the photo-responsive regulation of G-quadruplex (G4) structures for therapeutic or nanotechnological applications. This introductory chapter aims to give an overview of prior work in the field to date. To begin, a brief introduction to nucleic acid chemistry is provided, highlighting the features of these molecules that are crucial to their functional roles, especially their ability to form a wide range of secondary structures. This allows for a more detailed exploration of G4 nucleic acids, the focus of this thesis. The general features of G4 self-assemblies are first considered, and their structural polymorphism is surveyed in greater detail. Next, the relevance of these architectures in biological and nanotechnological applications is discussed, particularly the opportunities afforded by regulating these oligonucleotide-based structures with small molecule ligands. A focus is placed on photoresponsive G-quadruplex systems, owing to the specific advantages of light as a stimulus for the regulation of biomolecules, and previous efforts towards the development of such systems are discussed. This survey identifies both existing progress towards the development of photoresponsive G4 ligands and highlights the limitations in the current state-of-the-art that justify the new research described herein.

1.1 General nucleic acid chemistry

Nucleic acids are well known for their ability to form a wide range of higher-order structures, a property that allows these molecules to fulfil a variety of important functions.[2] In biology, the DNA double helix provides the means of information storage for the whole of life.[3] Meanwhile, RNA is the basis of one of nature's most impressive machines, the ribosome, that so elegantly translates structure to function.[4] The following sections briefly explore the fundamentals of nucleic acid chemistry that are responsible for the diversity of structural and functional roles of these nucleic acid systems.

1.1.1 Watson-Crick base pairing: the classical double helix

The monomeric unit of nucleic acids is the nucleotide, a small molecule composed of a ribose sugar and a nitrogenous base (together termed a nucleoside) conjugated to a phosphate group.[3] In RNA, the ribose bears a hydroxyl group at the 2' position of the carbohydrate ring; in DNA, the sugar is deoxygenated at this position (as drawn in Figure 1.1a). The polymeric

structure arises from the joining of a chain of nucleotides by phosphodiester bonds formed from condensation of the phosphate on the 5' position of one nucleotide with the 3' hydroxyl group of the adjacent monomer (Figure 1.1a).

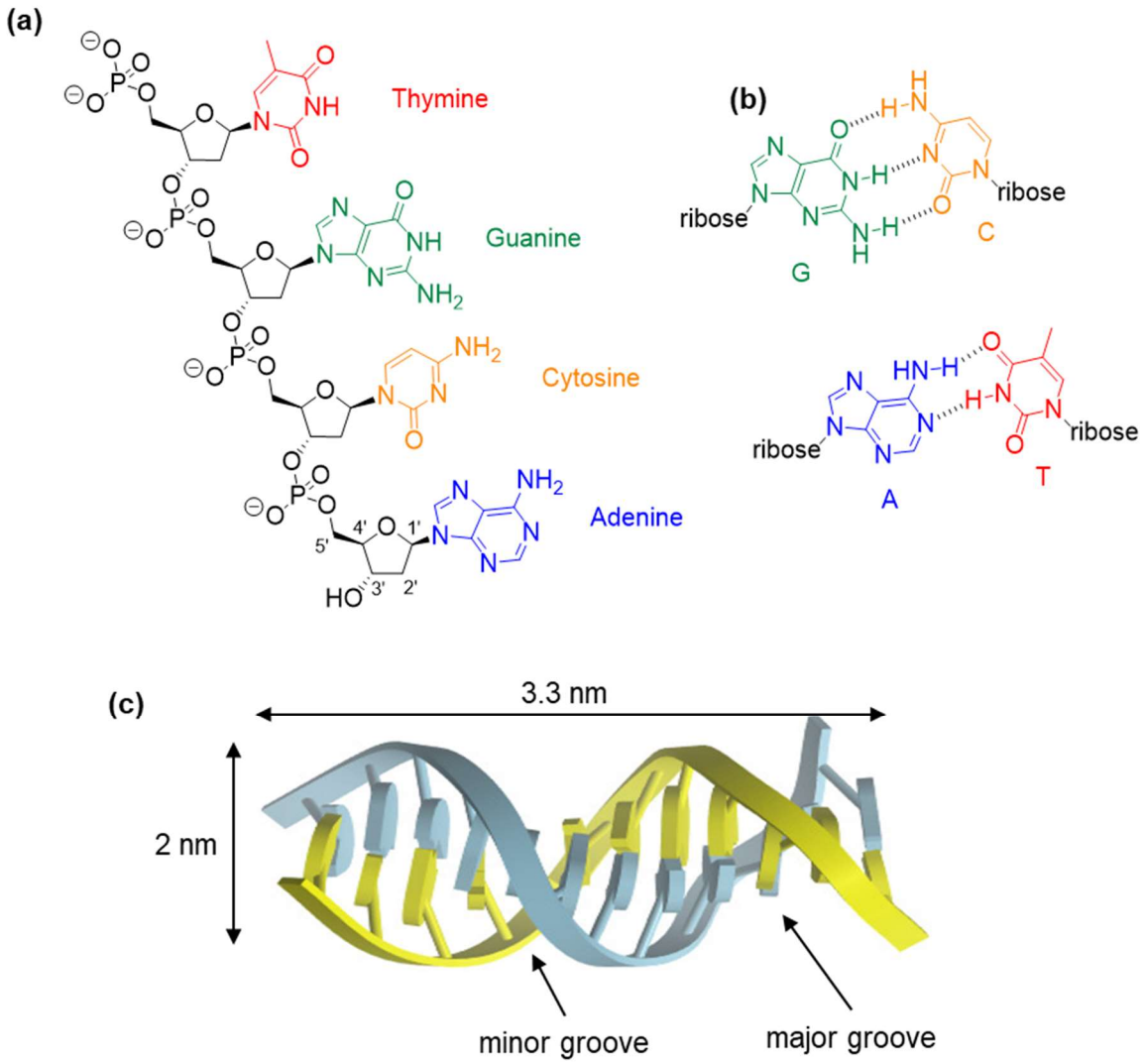


Figure 1.1: From nucleotides to B-DNA. (a) structure of a DNA oligonucleotide containing four constituent nucleotides chosen to represent the structures each of the four nitrogenous bases. The sugar carbon numbering is shown in the adenosine residue; (b) the formation of Watson-Crick GC and AT base pairs by complementary hydrogen bonding; (c) cartoon representation of the structure of B-DNA showing the double helix and associated dimensions. *Figure (c) reproduced from the PhD thesis of Steven T.G. Street,*[5] *generated from co-ordinates of PDB structure 1BNA.*[6]

Four nitrogenous bases occur naturally in DNA, namely guanine and adenosine (purine-derived bases) and the thymine and cytosine (pyrimidine-derived bases). The structures of the four bases are shown in Figure 1.1a. It is this variation that enables DNA to act as an information store, since information can be encoded in the sequence of bases that form the polynucleotide chain. A remarkable feature of these bases is the preferential association of adenosine with thymine and guanine with cytosine, which arises due to the complementary nature of the hydrogen bonding moieties in the two pairs (Figure 1b). The tendency for these base pairs to hybridise in this way is reflected in the classical "B-DNA" structure elicited in 1953 in which two complementary strands are wound together into a double helix.[7] The pitch and handedness of the DNA helix is determined by the subtle interplay of the conformation of the individual oligonucleotides in combination with the inherent chirality of the carbohydrate.[2,8] In B-DNA, the most common morphology found in biological settings, all nucleotides exhibit the *anti* conformation of the glycosidic bond angle (as presented in Figure 1a).[2] In combination with other conformational effects, such as C2'-endo puckering of the five-membered carbohydrate ring, this results in a right-handed helix with a pitch of 3.3 nm containing approximately 10 base pairs per turn (Figure 1c).[3] Meanwhile, the presence of *syn* glycosyl arrangements and alternative sugar ring conformations gives rise to further double-helical structures such as the right-handed A-form[9,10] and left-handed Z-form.[11,12] Whist still featuring classical Watson-Crick base pairing, the overall three-dimensional structures are very different (Figure 1.2).

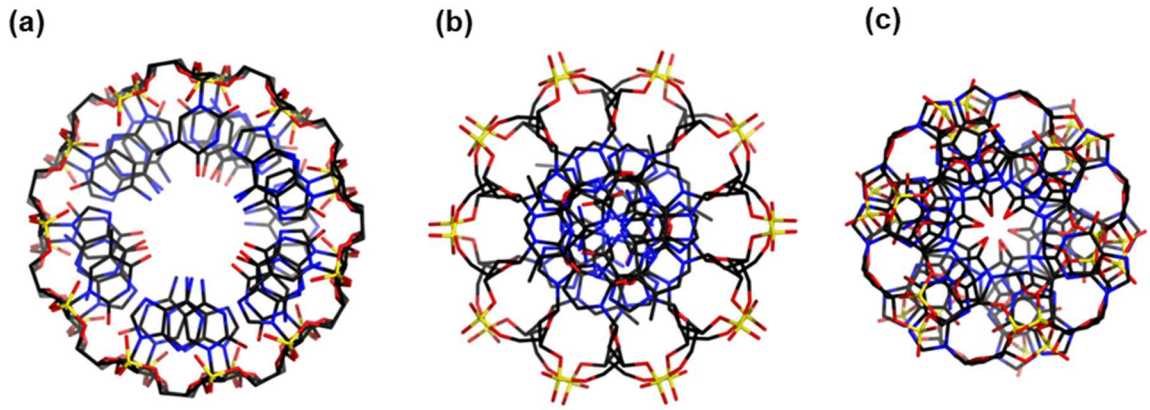


Figure 1.2: Top-down view of different DNA double-helical forms (a) A-DNA, (b) classical B-DNA and (c) Z-DNA. *Figure by Mauroesguerroto, distributed under a CC BY-SA 4.0 license.*

1.1.2 Hoogsteen base pairing: triplexes and beyond

Beyond the Watson-Crick arrangements observed in DNA double helices, further possibilities of base pairing exist. In 1956, Jerry Donohue proposed 24 possible pairings of the four DNA bases in which the geometrical arrangement allows at least two hydrogen bonds to form between homologous or heterologous pairs.[13] Though Donohue acknowledged that not all these pairings are likely to be relevant in biology, some evidence in support of additional modes of interaction was available at the time, such as the observation an A-A base pair in the crystal structure of adenine hydrochloride. More striking evidence was obtained in 1963 by Karst Hoogsteen, who reported that the 1:1 co-crystal structure of 1-methylthymine and 9-methyladenine revealed the two species hydrogen bonded in a non-Watson-Crick fashion involving the adenine N7 atom acting as the hydrogen bond donor, rather than the N1 atom as in the case of Watson-Crick base pairing.[14,15] The possibility for a given DNA base to form both Watson-Crick and Hoogsteen interactions allows the formation triad ensembles (Figure 1.3a), which enable the formation of triple-stranded DNA structures by hybridisation of a complementary third strand with the free Hoogsteen face of the bases accessible in the DNA major groove (Figure 1.3b).[16]

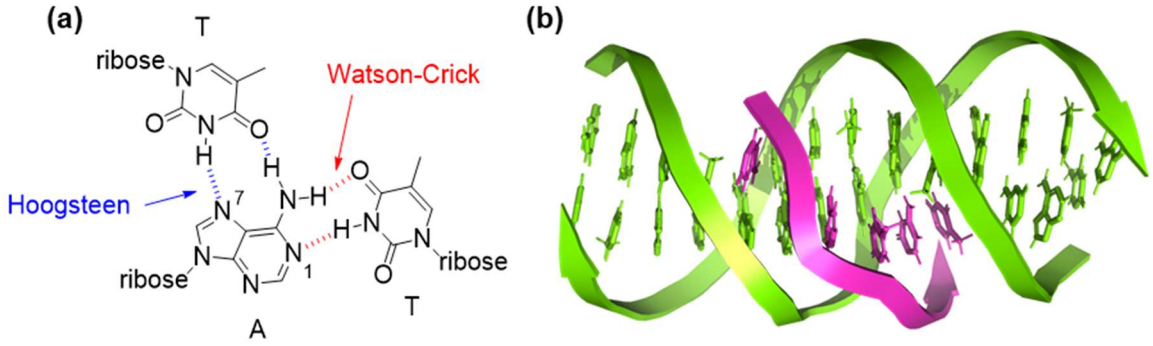


Figure 1.3. From Hoogsteen base pairing to DNA triplexes. (a) representation of Hoogsteen base pairing between thymidine and adenosine showing the involvement of N7 as a hydrogen bond donor. The Watson-Crick pairing base pairing is also shown. (b) cartoon representation of the structure of a DNA triplex consisting of a classical Watson-Crick (B-DNA) double helix (green) with a shorter oligomer (pink) bound in the major groove by association with the free Hoogsteen face of the DNA bases. *Figure (b) by Duboolan, distributed under a CC BY-SA 4.0 license.*

Despite the additional base pairing modes exemplified by these triple-stranded structures, the underlying canonical structure of B-DNA remains preserved, with the triplex-forming oligonucleotide in associating with the duplex almost in the manner of a ligand. However, Hoogsteen base pairing also offers further possibilities for the formation of nucleic acid secondary structures that do not resemble the classical double helix at all. In particular, the Hoogsteen face of guanine is particularly important in facilitating the self-assembly of nucleic acid sequences rich in this base to form a variety of supramolecular architectures. This topic is explored in more detail in the following section.

1.2 G4 nucleic acids: discovery and structural features

As early as 1910, it was reported by Bang that guanylic acid solutions spontaneously formed gels.[17] This observation is generally considered to be the earliest hint of the propensity of guanine to form supramolecular assemblies but the methods required to probe the nature of these structures in detail were not yet available. In 1962, Gellert and colleagues disclosed a study of the optical properties and X-ray diffraction patterns of these materials.[18] Through analysis of the diffraction patterns, the team were able demonstrate the aggregates were helical, ordered four-stranded structures. They therefore postulated the G-tetrad (Figure 1.4a) as the unit of assembly, in which four guanine residues are held together by complementary hydrogen bonding stabilised by co-ordination to a central metal cation. Self-association of many such tetrads by π- π stacking was proposed to form the helical gel-like superstructures (Figure 1.4b).

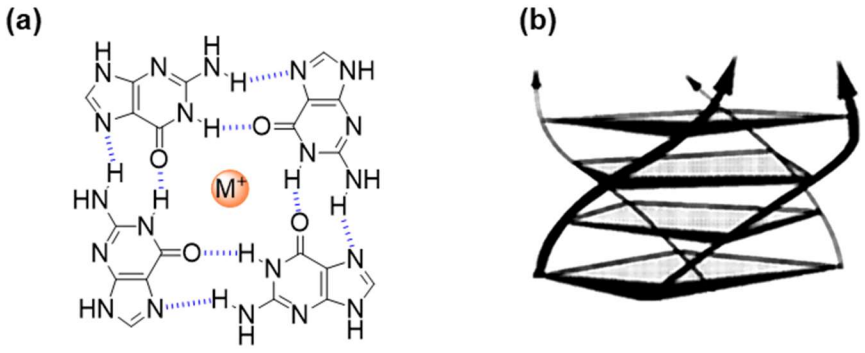


Figure 1.4. Model of self-assembly of guanylic acid proposed by Gellert and co-workers. (a) a G-tetrad formed from the Hoogsteen base pairing between four guanine residues; (b) proposed helical structure of tetrad self-assembly by π-π stacking. *Figure (b) used with permission, © National Academy of Sciences*.[19]

As can be seen from inspection of the proposed G-tetrad structure, both the Watson-Crick and Hoogsteen faces of the guanine bases are required to from the G-tetrad arrangement though, at the time, it is unlikely that Gellert and colleagues were aware of Hoogsteen's own discoveries. Gellert's observations clearly pointed to the potential relevance of guanine self-assembly in regulating the secondary structure of DNA in biology. Indeed, in the 1970s, similar diffraction studies on polyguanylic acid were reported, indicating that these species also form four-stranded helical structures in an analogous manner to the monomeric species.[20,21] However, it was not until 1988 that evidence for the self-assembly of biologically-relevant oligonucleotides was obtained by Henderson, Blackburn and colleagues.[22] Through NMR studies, the group observed that human telomeric DNA (discussed in more detail in Section 1.3.3.1) formed intramolecular structures, stabilised by non-canonical G-G base pairing, at low temperatures. Shortly afterwards, Sen and Gilbert observed that guanine-rich oligonucleotides spontaneously self-associated into four-stranded, tetrameric structures in monovalent salt buffer.[23] In light of their results, and those of Henderson and colleagues, Sen and Gilbert proposed that telomeric sequences may spontaneously form four-stranded structures under physiological conditions. These seminal reports led an intensive flurry of activity in the early 1990s that aimed to observe these structures more directly. Indeed, several X-ray diffraction[24] and NMR-based[25–27] studies at this time unequivocally demonstrated that four-stranded G-rich DNA, now referred to as the G-quadruplex, can form *in vitro*, though much work remained to further interrogate its biological relevance.

These structural studies on model G4-forming oligonucleotides revealed that whilst the G-tetrad is the common unit of assembly, significant variety exists in the way the nucleic acid strands can arrange themselves to form this network (Figures 1.5 and 1.6). Whilst the parallel tetramolecular quadruplexes postulated by Sen and Gilbert were observed experimentally by Wang and Patel (Figure 1.5a),[28,29] Rich and co-workers found that longer sequences containing intervening "loop" residues between the G-tracts allow the possibly fold-back arrangements, with subsequent dimerisation of two identical residues to form a bimolecular quadruplex (Figure 1.5b).[30] In these structures, alternate guanines are found to adopt the *syn* conformation to facilitate the necessary antiparallel arrangement between the strands, rather than the all-*anti* arrangements found in the parallel G-quadruplexes. Finally, longer sequences containing four G-tracts separated by loop-forming residues can fold to form unimolecular quadruplexes (Figure 1.5c), first observed for the thrombin-binding aptamer ($G_2T_2G_2TGTG_2T_2G_2$) in potassium-rich solution.[31] Shortly afterwards, a landmark study by Wang and Patel demonstrated the formation of an intramolecular G4 formed by the natural human telomeric sequence ($AG_3T_2AG_3T_2AG_3T_2AG_3T_2A$) in sodium-rich conditions.[27]

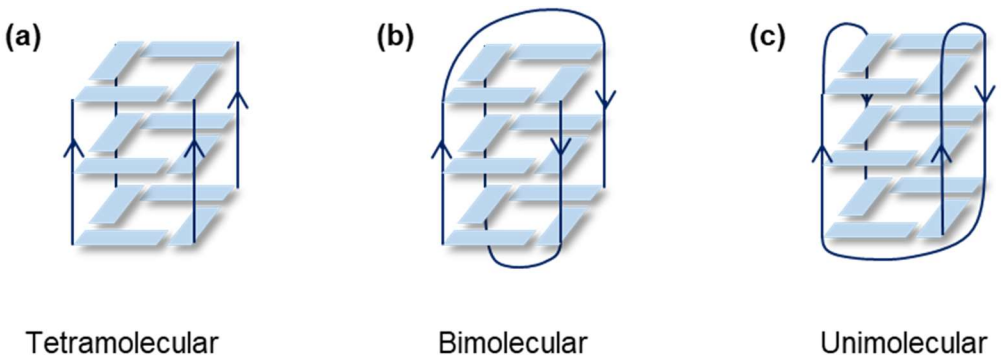


Figure 1.5. Schematic representation of quadruplex structures formed from different numbers of nucleic acid strands: (a) termolecular G-quadruplex (b) bimolecular G-quadruplex (c) unimolecular G-quadruplex.

Subsequent studies on unimolecular G-quadruplexes, particularly human telomeric DNA models, reveal a significant number of possible structures this sequence can form depending on the folding conditions.[32] The difference in folded arrangements give rise to distinctive features of each topology, in particular different loop types (diagonal, lateral and propeller/chain reversal), groove sizes and relative orientations (parallel/antiparallel) of the adjacent strands (Figure 1.6).[33] The structure adopted depends on a variety of factors, such as the precise sequence in question, the identity of the metal ion, and the presence or absence of crowding agents.[34,35] For example, the solution structure of telomeric DNA determined by Wang and Patel forms an antiparallel arrangement (known as the “basket form”), connected by two lateral loops and one diagonal loop (Figures 1.6a and 1.6b).[27] However, when the same sequence is crystallised in the presence of potassium ions, an all-parallel conformation is found where the strands are connected by chain-reversal (propeller) loops (Figures 1.6c and 1.6d).[36] The structures of several related G4 sequences have also been solved, revealing further possibilities of folding topology, such as the (3+1) hybrid form, containing a combination of chain-reversal and lateral loops (Figure 1.6e and 1.6f).[37]

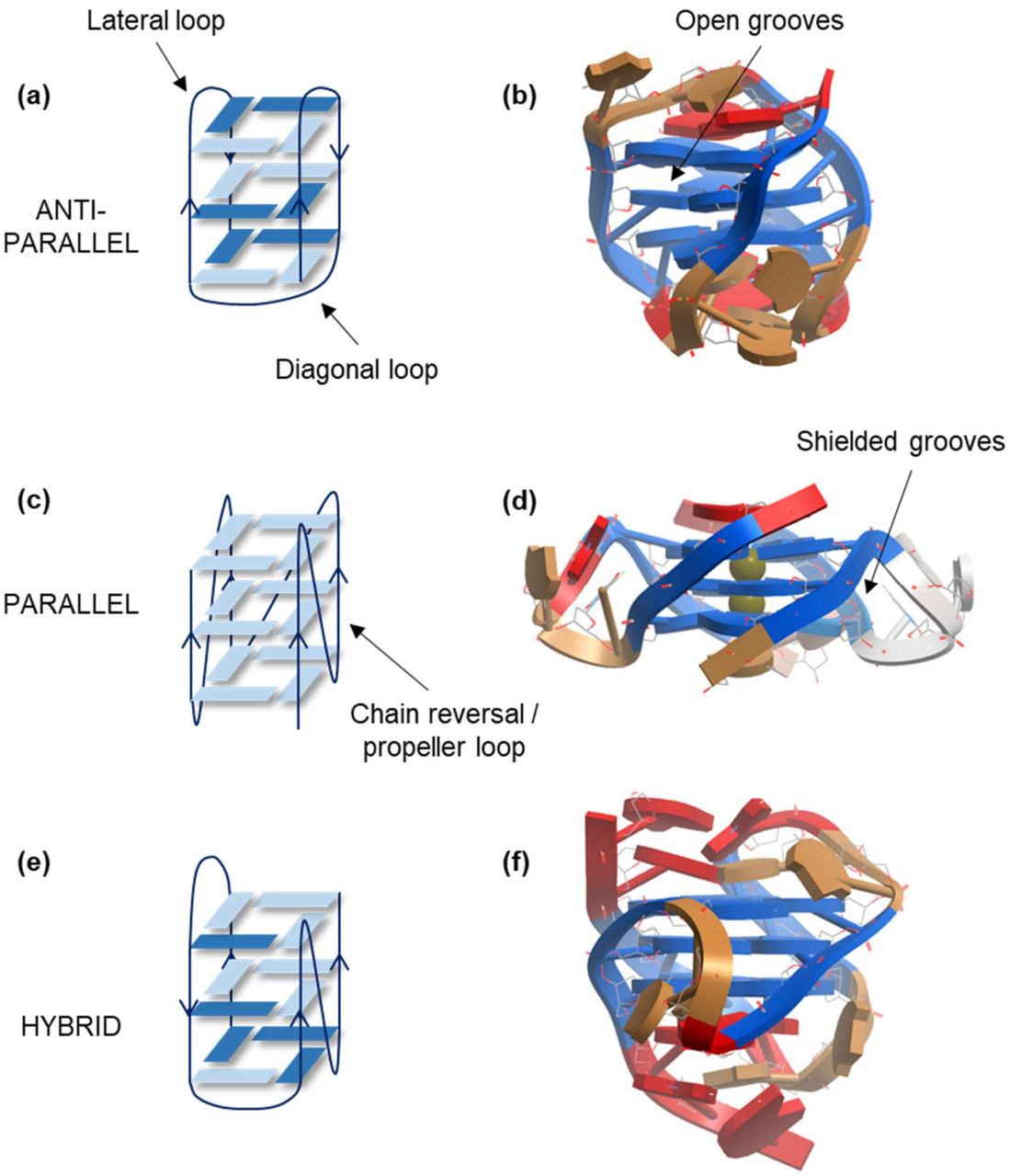


Figure 1.6: Schematic illustrations (a,c,e) and experimentally-determined structures (b,d,f) of different polymorphs of the human telomeric G-quadruplex sequence: (a)-(b) antiparallel G4 formed from two lateral loops and one diagonal loop (PDB ID: 143D);[27] (c)-(d) parallel G4 formed by chain-reversal loops (PDB ID: 1K8P);[36] (e)-(f) hybrid G4 formed from two lateral loops and one chain-reversal loop (PDB ID: 2AQY).[38] In (a,c,e) the strand directionality is indicated with arrows and the guanine residues are colour coded dark blue (*syn* conformation) and light blue (*anti* conformation). *Experimental structures reproduced from the PhD thesis of Steven T.G. Street,[5] generated from co-ordinates of the respective PDB structures.*

1.3 Control of G4 structure: relevance and applications

The discovery that G-quadruplex structures form in physiologically-relevant conditions generated huge interest across the scientific community in the wider relevance of these structures. The potential biological roles of G-quadruplex DNA have been, and continue to be, investigated in detail. G-quadruplexes have been observed in human cells through antibody staining[39] and *in vivo* NMR techniques,[40] and evidence implicating G-quadruplexes in a large number of regulatory roles in biological processes has been obtained.[41,42] Because of their biological relevance and their distinct structural features compared to B-DNA, G-quadruplexes have received significant interest as therapeutic targets, since their unique topography can be targeted specifically with small molecule ligands.[33,43–45] Meanwhile, the potential of G-quadruplexes to regulate the formation of nucleic acid architectures has garnered interest as a means to exploit the formation of these nanostructures as the basis of DNA-derived nanodevices.[46,47] This section briefly introduces these two research areas.

1.3.1 G4 as a therapeutic target

The level of interest in G4 partly stems from its potential as a therapeutic target. Indeed, G4-forming DNA and RNA sequences are not confined to the human genome but are also observed in a host of other organisms including plants,[48] bacteria,[49] viruses[50,51] and parasites.[52–54] Several bioinformatics algorithms have been developed to mine genomic sequences for putative quadruplex-forming motifs comprising G-tracts and loop residues of appropriate length.[55] Earlier tools, such as Quadparser, developed by Huppert and Balasubramanian,[56] have more recently been developed into more sophisticated algorithms, for example G4Hunter (developed by Mergny and co-workers) which permit more flexibility in the definition of putative G4-forming sequences to reflect the diversity in G4 structures in encountered to date.[57] These computational approaches indicate the prevalence of potential G4-forming sequences across the genomes of many species, suggesting many potential opportunities to exploit G4 for the development of new therapeutic strategies. Much of the excitement around targeting the G-quadruplex in therapy arises from the possibility to specifically target these structures over duplex DNA, based on the difference in size and shape of the relevant binding sites. For example, the surface area of a G-tetrad is approximately twice the size of a base pair of a classical Watson-Crick double helix.[58] Exploiting this selective mechanism of action is anticipated to reduce or eliminate the side-effects of traditional DNA-targeting drugs that induce general DNA damage and off-target toxicity.[59] Whilst this mechanism of action is designed to affect cancer cells most severely due to their abnormally high rate of proliferation[60] it also exerts deleterious effects on normal tissue due to the homogeneous nature of the majority of cellular DNA. Selective targeting of non-canonical DNA

structures to trigger anti-tumour action is therefore hypothesised to confer a therapeutic advantage.[33,44]

The following sections discuss the therapeutic relevance of G-quadruplexes and some illustrative examples of previously-developed G-quadruplex targeting agents.

1.3.1.1 Telomeric G-quadruplexes

Telomeres are found at the end of chromosomes and protect genetic material from damage, during normal cellular function and in cell replication.[61] The telomeric DNA is composed of repeating units of the sequence $(TTAGGG)_n$ and is found to be highly conserved across organisms.[62,63] In humans, telomeres are up to 15 kilobases in length and the DNA is associated with a range of proteins, which together are termed the shelterin complex.[64,65] A notable feature of telomeric DNA is that the 3'-end bears a single-stranded overhang of between 15 and 200 nucleotides. During cellular replication, DNA polymerase is unable to replicate the full length of this overhang, meaning the telomeres progressively shorted with each replication cycle.[66,67] Eventually, the shortening of the single-stranded overhang telomere reaches a point (known as the "Hayflick limit") which triggers a variety of process that lead to cell death.[68] Therefore, as well as protecting the genetic material, the telomere also regulates the cellular lifetime (Figure 1.7).[69]

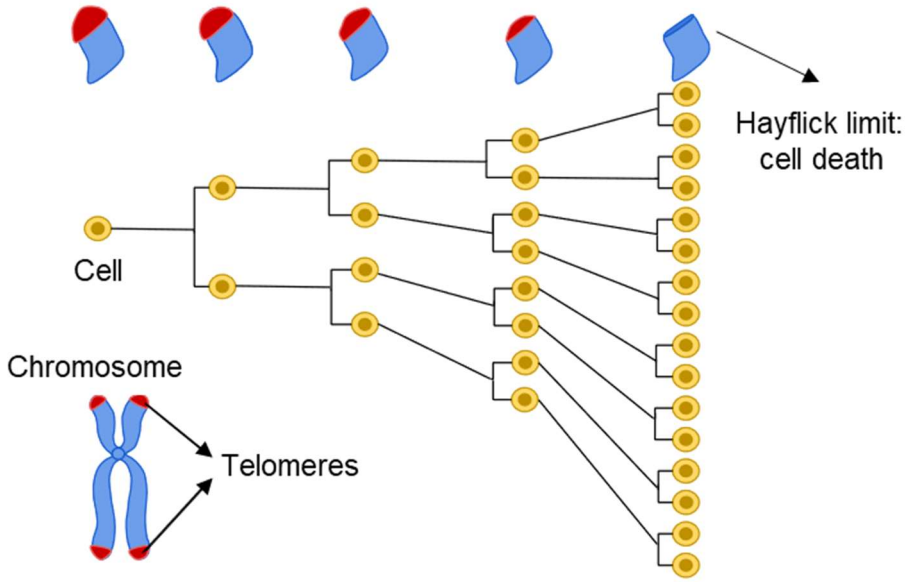
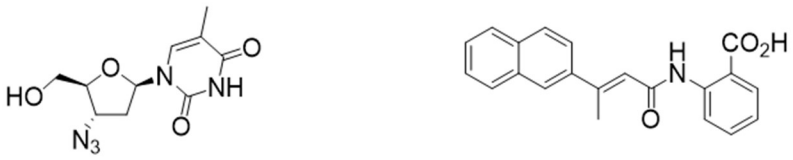


Figure 1.7: Illustration of the regulation of cellular lifetime by telomere shortening. With each cell division the telomere (represented as red) becomes shorter until the Hayflick limit is reached. *Figure by Azmitowski17, distributed under a CC BY-SA 4.0 license.*

A common feature of cancer cells is the upregulation of telomerase, a reverse transcriptase that maintains the length of telomeres by adding TTAGGG sequences to the 3' end of telomeric DNA.[70] The activity of telomerase is closely associated with cancer cell immortality, since the maintenance of telomere length leads to uncontrolled proliferation as the Hayflick limit is not reached.[71,72] Indeed, the enzyme is found to be up-regulated in 85-95% of cancers and comparatively non-expressed in the majority of non-cancerous cells.[73] Thus, inhibition of this enzyme receives significant attention as an anti-cancer strategy. Several approaches for the inhibition of telomerase have been pursued, including nucleotide-based reverse transcriptase chain-terminators such as the anti-HIV drug AZT[74] (**1**, Figure 1.8) and small-molecule inhibitors such as BIBR1532[75] (**2**, Figure 1.8). Though the latter compound proves effective in inhibiting telomerase (IC_{50} = 92 nM) *in vitro*, very long exposure times (of the order of 120 days) appear to be required before critical shortening of the telomere is achieved in order to promote senescence, questioning the utility of direct telomerase inhibition as a therapeutic strategy.



1 2

Figure 1.8: Examples of direct telomerase inhibitors: AZT (**1**) and BIBR2 (**2**).

As an alternative strategy of telomere targeting, the potential of exploiting the propensity of telomeric sequences to form G-quadruplexes was first ventured in a seminal study by Zahler and co-workers, reported in 1991.[76] The team demonstrated that the activity of telomerase is stalled when the DNA is folded into quadruplex form, prompting the notion that stabilisation of the G-quadruplex by small molecule ligands may indirectly inhibit telomerase. This observation led to the therapeutic hypothesis that induction of the telomeric G-quadruplex *in vivo* by a small molecule binding ligand may indirectly inhibit telomerase by preventing the binding of the enzyme to its DNA target (Figure 1.9). Proof of this concept was achieved by Hurley, Neidle and co-workers in 1997, who reported the first G-quadruplex stabilising ligand, an anthracene-9,10-dione (**3**, Figure 1.10), that displayed potent inhibitory activity (IC_{50} = 23 μM) of the enzyme.[77]

Following Hurley's report, significant attention has been devoted to the design of further G-quadruplex binding ligands as potential therapeutic agents, especially in light of further possible quadruplex targets identified across the genomes of humans and pathogens [44,78–84] Landmarks in G4 ligand development are discussed later (Section 1.3.1.4) in the context of general G4-ligand design principles.

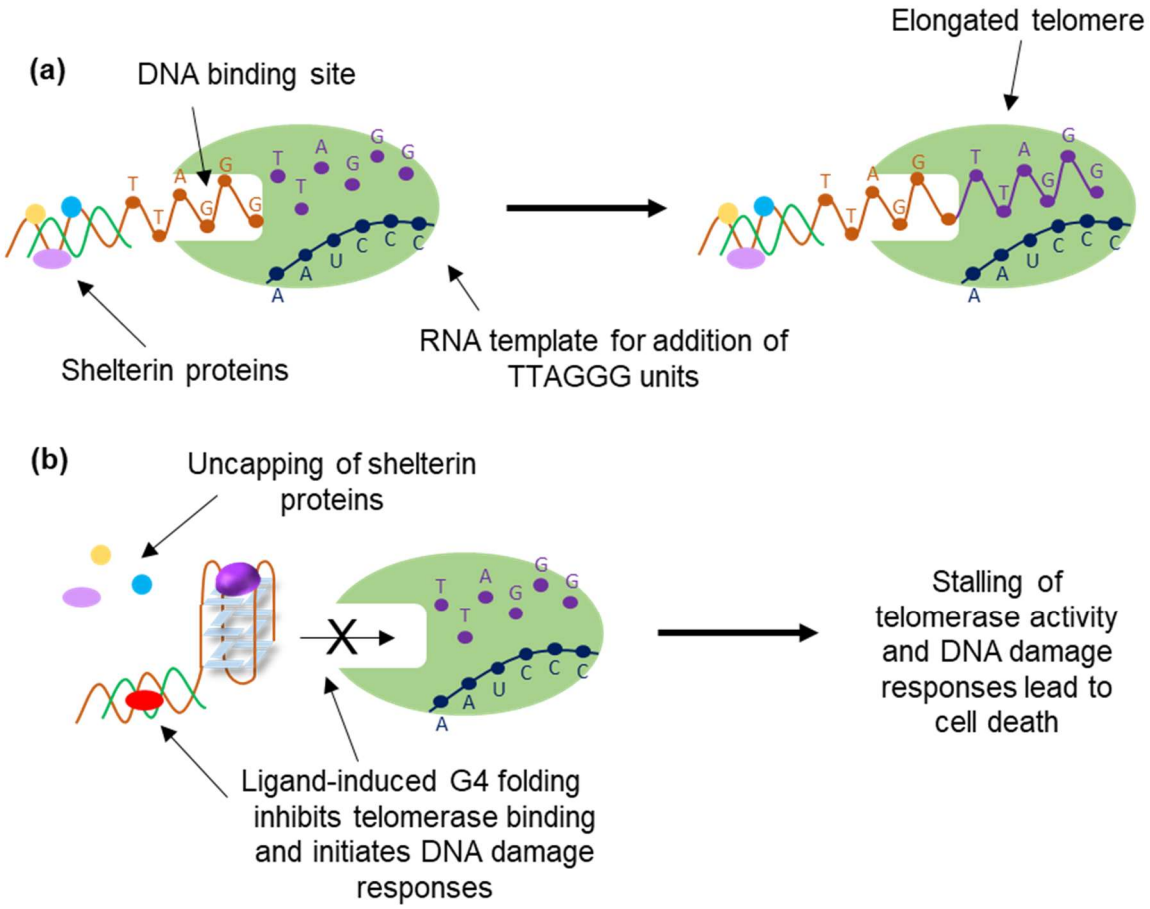
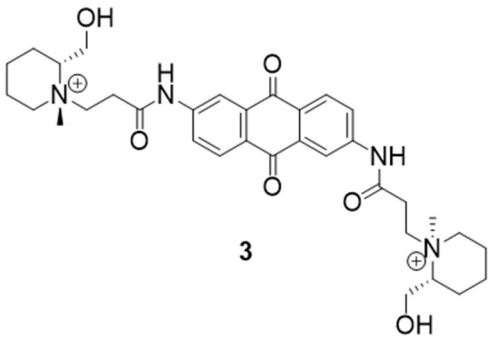


Figure 1.9: Therapeutic mechanism of telomerase inhibition by G-quadruplex ligands (a) telomerase acts on single-stranded DNA; (b) stabilisation of G-quadruplex by ligand prevents telomerase binding, triggers disruption of the shelterin complex and induces DNA damage responses.



3

Figure 1.10: The first G-quadruplex ligand shown to indirectly inhibit the activity of telomerase, reported by Hurley, Neidle and co-workers in 1997.

Though G4-binding compounds were initially designed based on a therapeutic hypothesis of telomerase inhibition, the antiproliferative effects of telomeric G4 binding ligands often appear to be more acute than those observed with direct inhibitors of telomerase such as BIBR1532, suggesting greater clinical utility based on a more rapid response, but also pointing to additional mechanisms of action than indirect telomerase inhibition.[44] Indeed, it has subsequently been shown that the stabilisation of the telomeric G4 triggers the uncapping of key shelterin proteins from the telomere,[85,86] leading to short-term DNA damage responses which induce telomeric dysfunction and associated cellular senescence and apoptosis on much shorter timescales than required for Hayflick-related cell death.[87] It has been ventured that the observed selectivity for tumoral cells may be due to an absence of defence mechanisms against G4-folding in cancerous cells, such as the helicase enzymes capable of unwinding the telomeric quadruplex.[44,88,89] However, the true mechanism of action of G4 targeting compounds is difficult to establish with certainty, not least because the same ligand may also target G-quadruplexes at locations other than telomeres (such as those in promoters sequences) that may be also responsible for more short-term destructive effects. These non-telomeric G-quadruplexes are briefly discussed in the following section.

1.3.1.2 G-quadruplexes in oncogene promoters

Beyond the telomere, bioinformatics approaches have revealed an abundance of putative quadruplex forming sequences across the human genome.[57] Experimentally, sequencing studies by Balasubramanian and co-workers identified 716,310 putative G4-forming sequences in genomic DNA[90] and, more recently, a chromatin immunoprecipitation strategy (ChIP-seq) by the same laboratory observed the formation of approximately 10,000 G4 structures in endogenous human chromatin,[91] further strengthening the evidence for G4 in the regulation of genetic processes. Complementary experimental studies have validated the occurrence of many G-quadruplexes structures in the promotor regions of several genes associated with human cancers, such as the *c-myc*,[92] *BCL*-2,[93] *HRAS*,[94] *KRAS*,[95] c-*KIT*[96] genes, in addition to many others. Critically, G4-forming sequences are over-represented in oncogene promoters (genetic regions which are responsible for the activation of gene transcription), with 69% of these sequences found to contain a putative G4, versus a global occurrence of 43% across all promoter sequences.[97] In 2002, Hurley observed that the G-rich strand of the *c-myc* promoter was able to form G-quadruplex structures, and that the stability of this structure correlated with transcriptional activity.[92] Moreover, it was observed that the presence of a G-quadruplex stabilising compound, TMPyP4 (**4**, Figure 1.11) led to transcriptional suppression.

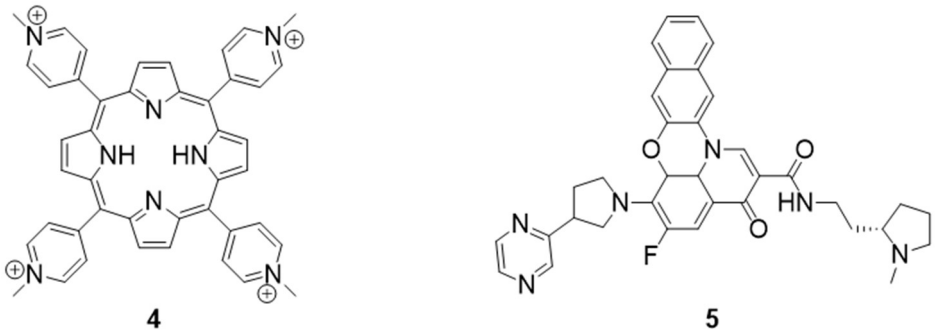


Figure 1.11: G-quadruplex binding molecules TMPyP4 **3** and quarfloxin **4** that target promoter G4.

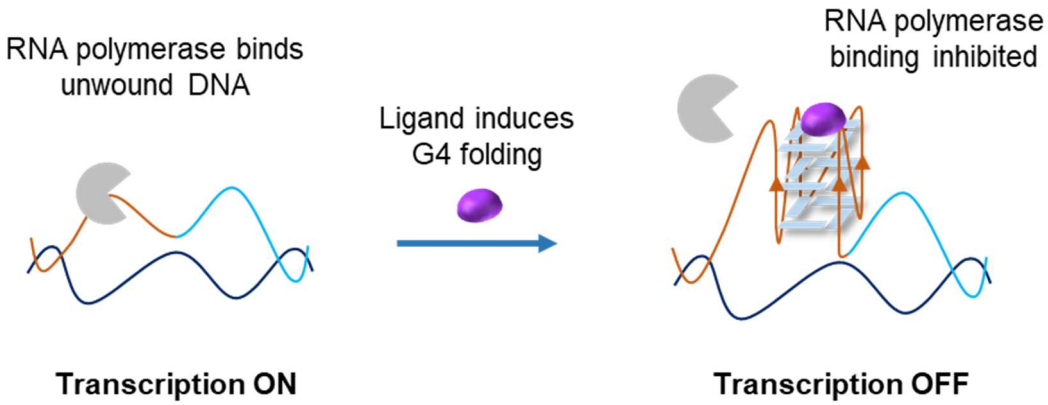


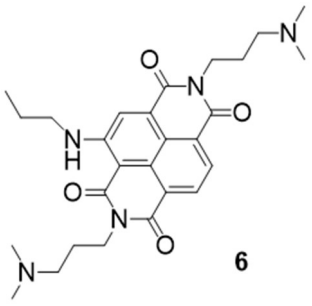
Figure 1.12: Proposed mechanism of oncogene suppression by targeting G4 DNA with small molecules. Induction of the G4 fold in the gene promoter by the stabilising ligand inhibits the binding of RNA polymerase and transcription factors leading to downstream silencing of the gene. *Figure adapted from the author's publication with permission © Wiley*.[1]

Based on these findings, it was proposed the stabilisation of the folded structure prevents the binding of transcription factors to the promoter leading to downstream silencing of the gene, and therefore targeting G-quadruplexes in promoter regions was introduced as a possible anticancer strategy (Figure 1.12). This concept ultimately led to the development of the compound quarfloxin (**5**, Figure 1.11) which was, until recently, the only G4 targeting compound to reach clinical trials.[45] Though this molecule is shown to bind strongly to *c-myc* (and other G-quadruplexes) *in vitro*, its biological mechanism of action was found to be more complex than that represented in Figure 1.12, instead displacing nucleolin from an RNA

quadruplex inside the cell nucleolus, which in turn acts as the *c-myc* G4 ligand inside the nucleoplasm leading to downstream apoptosis of cancer cells.

1.3.1.3 G-quadruplexes as antiparasitic targets

Though the vast majority of efforts on G4-targeting molecules focus on targeting G4 in the human genome as a potential anticancer strategy, attention is increasingly being focused on G4 sequences found in the genomes of pathogenic organisms such as bacteria[49] and viruses[50,51] as a means of developing much-needed efficacious treatments against these sources of infection. Recently, a bioinformatics search revealed the occurrence of G4 sequences in the genomes of the protozoan parasites *Trypanosoma brucei* and *Leishmania major*.[52] Both organisms contain frequent occurrences of the conserved telomeric sequence in addition to several further unique putative G-quadruplex forming sequences. For example, the EBR1 sequence d[$G_3CAG_5TGATG_4AG_2AGCCAG_3$] occurs 33 times in the *T. brucei* genome and was subsequently demonstrated to form a stable quadruplex under physiological conditions in biophysical studies.[52] The sequence predominately occurs in genomic regions coding for several proteins including a cysteine peptidase and a purine transporter. G4s therefore present a potential opportunity as a target for novel antiparasitic therapies, for which there is an urgent need for further development.[98] For example, the *T. brucei* parasite, responsible for the Human African Trypanosomiasis (HAT) disease, endangers 69 million people across Sub-Saharan Africa,[99,100] and existing therapies (such as pentavalent antimonial compounds) suffer from severe limitations including severe side effects[101] and emerging drug resistance[102] in the parasitic strains.



6

Figure 1.13: Example of a naphthalene diimide G4-targeting ligand that induces selective toxicity in *T. brucei* and *L. major* parasites over non-tumoral mammalian cells.

The handful of G-quadruplex binding compounds explored for antiparasitic activity to date are primarily naphthalene diimide derivatives.[52,103] For example, the G-quadruplex targeting compound **6** (Figure 1.13) displays low nanomolar (EC_{50} = 9 nM) toxicity to *T. brucei* parasites

whist remaining approximately 40-fold less toxic to non-tumoral mammalian cells. The same compound is also toxic to *L. major* (EC_{50} = 34 nM). Though the activities observed in these early studies are promising, the identification of further DNA-binding chemotypes capable of exerting selective antiparasitic activity is of critical relevance to exploring this therapeutic hypothesis. Furthermore, the detailed mechanism of action of these compounds remains to be fully elucidated.

1.3.1.4 G4 ligand design and screening

The vast majority of G4 ligands employ rigid aromatic heterocyclic frameworks that are designed to π-stack onto the large surface of the G-tetrads in preference to intercalating the base pairs of DNA duplexes.[79,104] Cationic character (for aqueous solubility and to facilitate electrostatic interactions with negatively charged nucleic acids) is often introduced by conjugation of these cores to basic amino residues (which are protonated to the ammonium form at physiological pH), to quaternised ammonium or pyridinium moieties, or to transition metal cations.[1,44] Ligands may also target the G4 through other binding modes, such as to the loops[105] or grooves[106] as well as the planar tetrads.

Standard biophysical techniques are used to probe the strength of small-molecule ligand binding to G-quadruplexes, such as spectroscopic or calorimetric titration experiments.[107] DNA melting assays are frequently employed to examine the ability of molecules to stabilise the folded G4 structure to thermal denaturation which also provide an indication of ligand binding.[108] The binding modes of ligands to the G4 structures may be elucidated using NMR spectroscopic techniques[109] and X-ray crystallography[110] whilst circular dichroism spectroscopy is a powerful tool to characterise the effect of the ligand on the G-quadruplex topology, since the different topological arrangements give rise to distinctive spectral signatures.[111] Several of these techniques were employed during the research reported in this thesis, and they are introduced in more detail in the relevant results chapters. Meanwhile, validation of a ligand's mode of action requires specific *in vitro* assays, such as the TRAP (telomeric repeat amplification protocol) assay, which allows the ability of a given molecule to inhibit the telomerase to be quantified,[112] or gene transcription assays in the case of ligands designed to target promoter quadruplexes.[92]

Some well-studied G-quadruplex ligands are displayed in Figure 1.14 to illustrate ligand design principles. Following the discovery that anthracene-9,10-dione ligand **3** (Figure 1.10) inhibits telomerase activity by stabilising G-quadruplex DNA, significant efforts were undertaken to optimise the quadruplex binding properties of this scaffold in pursuit of a therapeutic agent.[113,114] However, it was later identified that the relatively low

quadruplex/duplex selectivity of these derivatives was unlikely to be sufficient for further biological application. Following iterative design by Neidle and co-workers, modifying the aromatic core[115,116] and groove-binding side-chains,[117] it was found that the addition of a third side-chain substituent to an acridine scaffold conferred much higher G-quadruplex selectivity than observed with the bis-substituted analogues. The emerging lead compound from these studies, BRACO-19 (**7**, Figure 1.14) confers a 27 °C thermal stabilisation to folded quadruplex DNA (measured by a fluorescence-based melting assay)[104] and the compound was

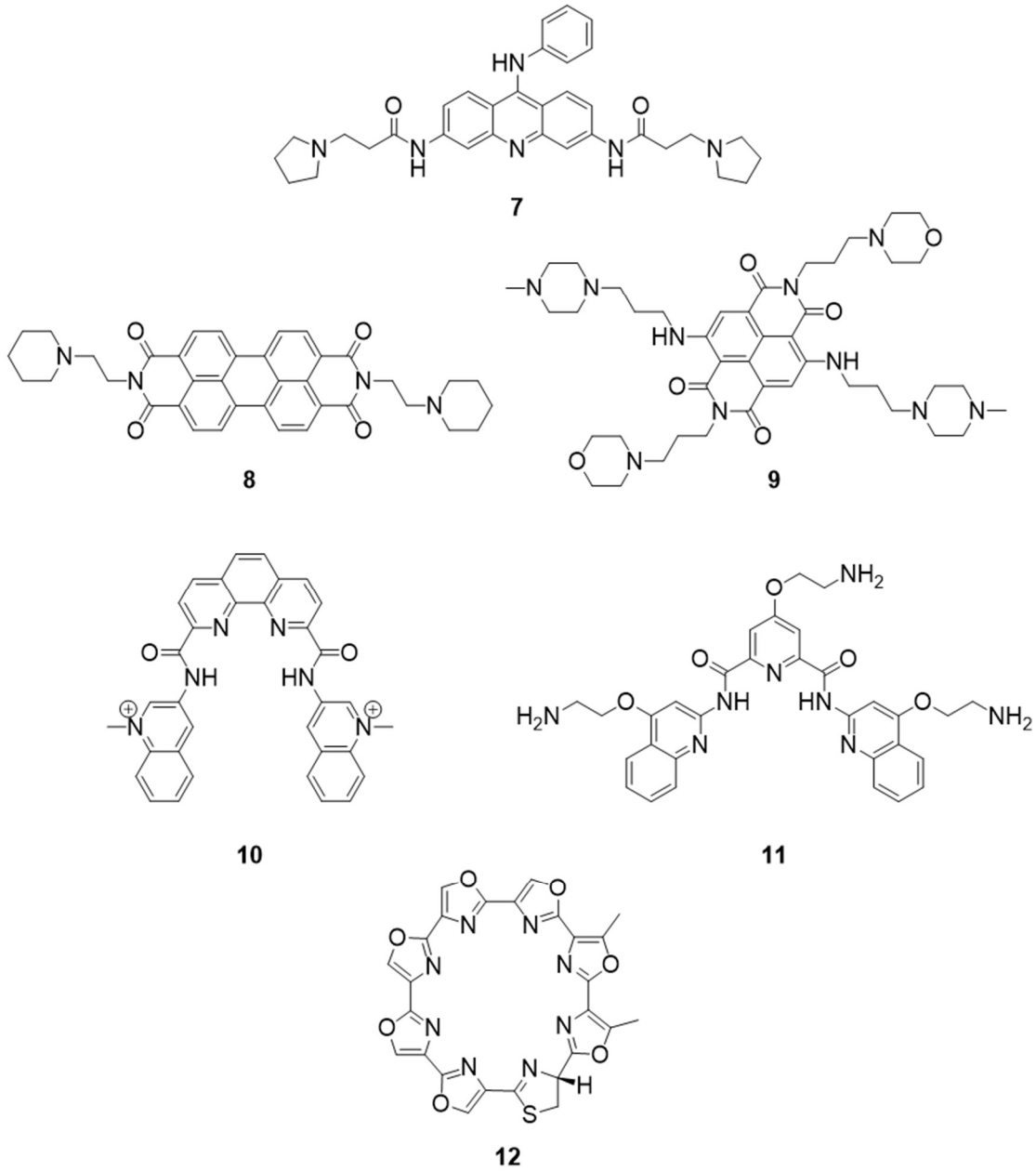


Figure 1.14: Examples of previously-developed G4 ligands.

subsequently shown to inhibit tumour growth with low micromolar efficacy, with evidence that

the toxicity results from telomerase inhibition.[118] X-ray crystallography studies indeed demonstrated the stacking of the aromatic core with the planar G-tetrads of human telomeric DNA, whilst the basic side chains targeting the G4 grooves.[110]

Earlier efforts from Hurley's group focused on the perylene diimide motif, leading to the development of PIPER (**8**, Figure 1.8).[119] PIPER shows particular selectivity for quadruplex DNA when in an aggregated state at pH 8.5.[120] However, little selectivity between DNA structures is observed at pH 7 and the very large, flat nature of the central core would be unlikely to confer favourable pharmacokinetic properties *in vivo*.[121] This perhaps led Neidle and co-workers to focus their studies on the more compact naphthalene diimide (NDI) scaffold. These compounds have been extensively developed by both Neidle's group[122] and others[123–126] in recent years, leading to some very promising lead compounds such as MM41 (**9**, Figure 1.8). Molecular modelling studies suggest that the tetrasubstitution pattern of these ligands is a key factor in achieving high G-quadruplex binding selectivity due to simultaneous recognition of four quadruplex grooves.[127] However, studies in our group have recently shown that divalent NDIs also display good selectivity for human telomeric DNA in potassium buffer, suggesting a large number of groove-binding motifs is not necessarily essential for quadruplex/duplex selectivity.[124] MM41 is currently advancing towards potential therapeutic application and has recently shown anti-tumour activity against pancreatic cancer in mice.[128]

Nitrogen-containing heteroaromatic groups appear to be rather privileged scaffolds in G-quadruplex ligand development. TMPyP4 (**4**, Figure 1.11), containing four *N*-methylpyridinium moieties induces high thermal stabilisation (17 °C)[104] in G4 structures but, likely by virtue of its high cationic charge, it is non-selective between G-quadruplex and duplex structures.[129] However, a significant development in G-quadruplex ligand design was the observation that bis-quinolinium functionalised phenanthroline derivatives are highly effective scaffolds for targeting the four-stranded structure.[130] Indeed, PhenDC3 (**10**, Figure 1.14) is one of the most potent and selective G-quadruplex ligands reported to date. In this case, the crescent-shaped molecular confirmation promoted by intramolecular hydrogen bonding renders the surface area of the molecule almost perfect for overlap with square-planar G-tetrads. This is reflected in the excellent stabilisation of G-quadruplex structures (up to 30 °C at 1 μM ligand concentration) whilst the stability of duplex DNA is unaffected. Moreover, the ligand-induced G4 stabilisation is retained even in the presence of high concentrations of duplex competitor, indicative of the discrimination between the two structures. Another benchmark G-quadruplex ligand is pyridostatin (**11**, Figure 1.14) which also contains the quinoline moiety though, in this case, as the free base form.[131] The G4 affinity in this case is therefore afforded by the basic amine residues in the side chains when protonated at physiological pH, but also by the molecular shape of the aromatic scaffold which, like PhenDC3, is crescent-like owing to

intramolecular hydrogen bonding between the pyridine nitrogen lone pair and amide NH group.

The natural product telomestatin (**12**, Figure 1.14) was first isolated *from Streptomyces annulatus* in 2001.[132] This compound induces high thermal stabilisation of quadruplex DNA of 24 °C.[104] Interestingly, these properties arise in spite of the lack of side-chain groove-binding motifs that are characteristic of many other G-quadruplex ligands. Instead, the activity of telomestatin activity results from the extensive overlap of its delocalised π-electron system with the G-tetrad, as demonstrated by the solution structure of the molecule in complex with telomeric DNA determined by NMR spectroscopy.[133] These characteristics are considered to result in the exceptionally high (70-fold) selectivity for quadruplex DNA, with the cyclic shape being unsuitable for interaction with DNA duplex. Telomestatin showed promise in early biological studies, demonstrating ability to induce cell death in leukaemia cells by telomerase inhibition and telomere shortening.[134] However the difficulty in obtaining telomestatin by natural and synthetic methods,[135] as well as its poor aqueous solubility,[79] have rendered it unsuitable for clinical development.

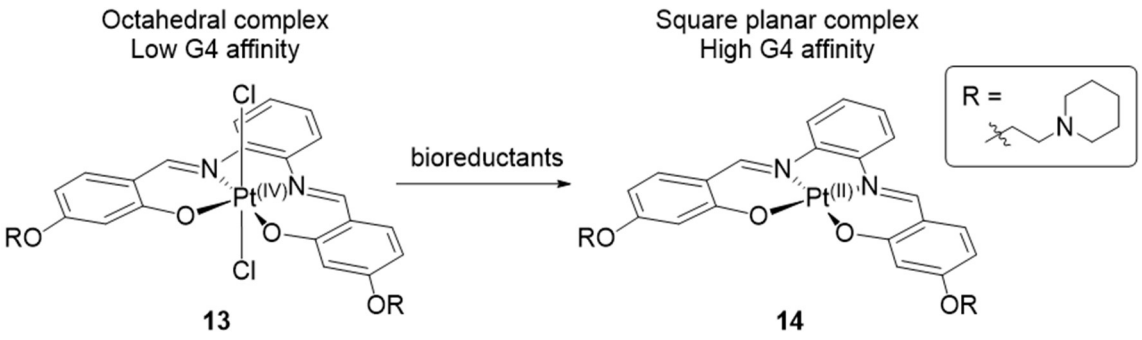


Figure 1.15: A redox-activated G4 ligand**.**

Metal complexes have also been employed as G4 ligands, since the square planar coordination compounds are well-suited to forming stacking interactions with G-tetrads.[136] This property allows the engineering of stimuli-responsive ligands based on the redox properties of the metal centre, as demonstrated recently by Vilar and co-workers through the design of platinum complex **13** (Figure 1.15).[137] This octahedral Pt(IV) complex is a poor G4 binder, but is readily reduced to the G4-binding square planar Pt(II) complex **14** in the presence of bioreductants, such as glutathione or ascorbic acid, suggesting that the compound may serve as a prodrug to target the hypoxic environment of tumour cells.

1.3.2 G4 as the basis of functional materials

Though G-quadruplexes have clear potential as therapeutic targets by virtue of the ubiquity of nucleic acids amongst all living organisms, this does not restrict the exploitation of G-quadruplexes solely to biological applications. In particular, the defined polymorphism of these structures points to a variety of applications in the development of nanodevices, since interconversion between G4 topologies, or between G4 and unfolded structures, may be deployed to perform mechanical work at the molecular level.[47,138] Alternatively, the switching of G4 between distinct conformations may alter its interactions with other components of a system, allowing an input (which triggers a change in G4 conformation) to be coupled to an output activated by the conformational change, such as the release of a molecular cargo in a delivery system[139] or the light-up of a fluorescent reporter.[140] Towards these applications, strategies that allow the responsive control of the G4 assembly are key to realising potential applications. This section briefly considers the major ways G4 structure can be reversibly configured with the aim of deploying these structures in functional materials.

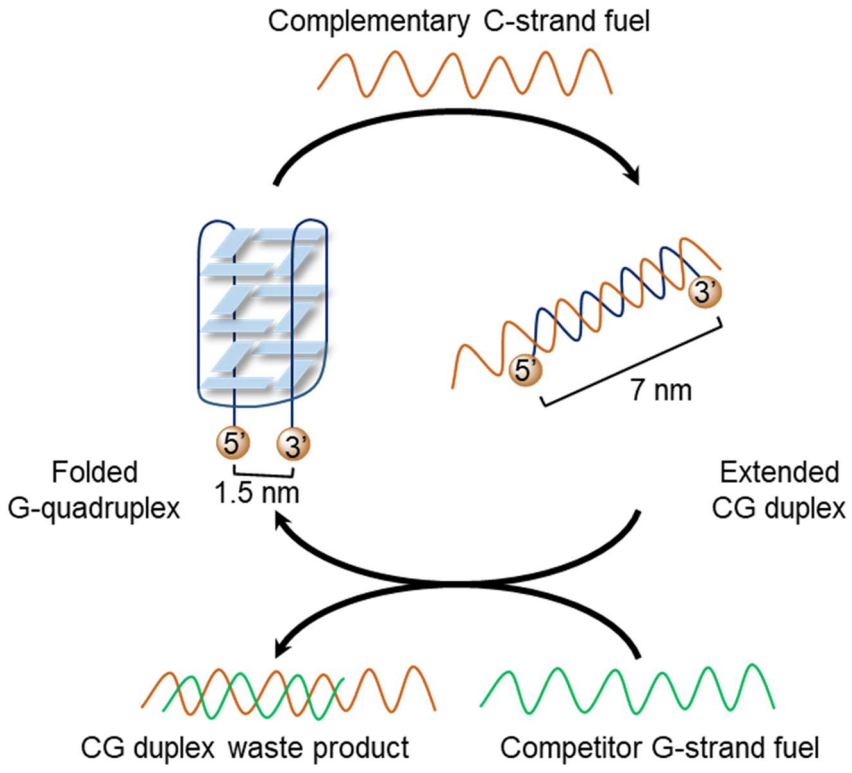


Figure 1.16: The G4-derived nanomachine developed by Alberti and Mergny. *Figure adapted from the author's publication with permission © Wiley*.[1]

1.3.2.1 Nucleic acid triggers

In 2003, Alberti and Mergny reported the first example of a G4-based nanomachine, which relies on a DNA-hybridisation strategy (Figure 1.16).[141] Addition of the complementary Watson-Crick strand to a folded G-quadruplex (i.e. a C-rich strand) triggers the unfolding of the quadruplex to allow hybridisation with the second strand, thus promoting the conversion of the four-stranded folded structure to an extended duplex. Indeed, the effective diameter of the oligonucleotide changes from 1.5 nm in the folded state to 7 nm in the single stranded form. The effect can then be reversed by addition of a further G-rich fuel, which outcompetes the G4-forming sequence for hybridisation with the C-strand, regenerating the folded G4 and a duplex DNA waste product. The reversible device could be switched over eleven cycles without observable fatigue by addition of appropriate complementary DNA strands. Such properties could possibly be exploited to position molecular cargo or exert force at the molecular level. Indeed, an application of this approach was later demonstrated by Simmel and co-workers, who developed a similar system based on the thrombin binding G4 aptamer.[139] The folded G4 form has high affinity for thrombin (a human blood-clotting enzyme), whist triggering of the unfolded form (through addition of a complementary oligonucleotide in the matter shown in Figure 1.16) caused the cargo to be released. However, in these strategies, the need for repeated additions of two different DNA fuels and the generation of the waste oligonucleotide products are potential drawbacks of the approach in terms of practical applications.

1.3.1.2 Metal ion triggers

Considerable success in exploiting G4 for functional applications has been achieved by exploiting the sensitivity of the structure to metal ions. In particular, dynamic control of quadruplex formation and dissociation can by readily achieved by controlling the concentration of free potassium ions, which trigger the folding of the four-stranded structure through stabilisation of the G-tetrads. Subsequent addition of a chelating agent, such as 18-crown-6, sequesters the metal ion causing unfolding of the G4 structure. Several applications of this metal-mediated conformational switch have been demonstrated.[46] For example, Willner and co-workers have reported DNA-based hydrogels where the stiffness of the material can be reversibly controlled by the potassium/18-crown-6 mediated crosslinking of the constituent strands into G-quadruplexes (Figure 1.17a).[142]

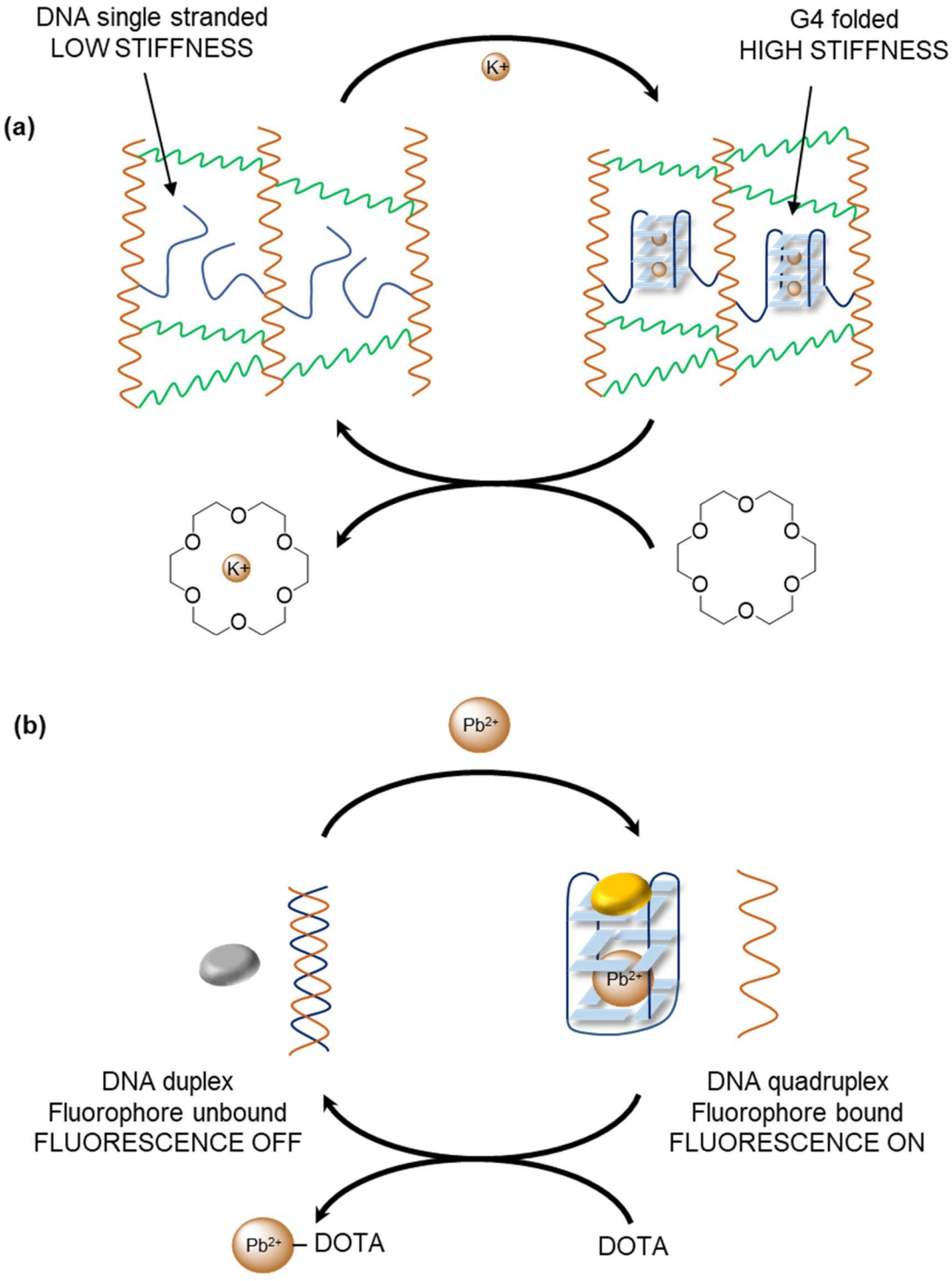


Figure 1.17: functional applications of controlling G4 assembly with metal ions and chelators (a) switching of hydrogel stiffness by alternate dosing of K^+ and crown ether; (b) a Pb^{2+} sensor based on the activation of fluorescence on Pb^{2+}-induced G4 formation.

In a different approach, Wang and co-workers exploited the Pb^{2+} formation of a G-quadruplex structure to act as a sensor for the metal ion itself (Figure 1.17b).[140] In this strategy, the metal-mediated stabilisation of the G4 structure provides a binding site for a fluorophore which becomes emissive upon binding, thereby indicating the presence of the metal ion. In a similar way to the potassium-responsive hydrogels developed by Willner, the device may be recycled by addition of a suitable chelating agent, in this case DOTA, a cyclic tetra-amine.

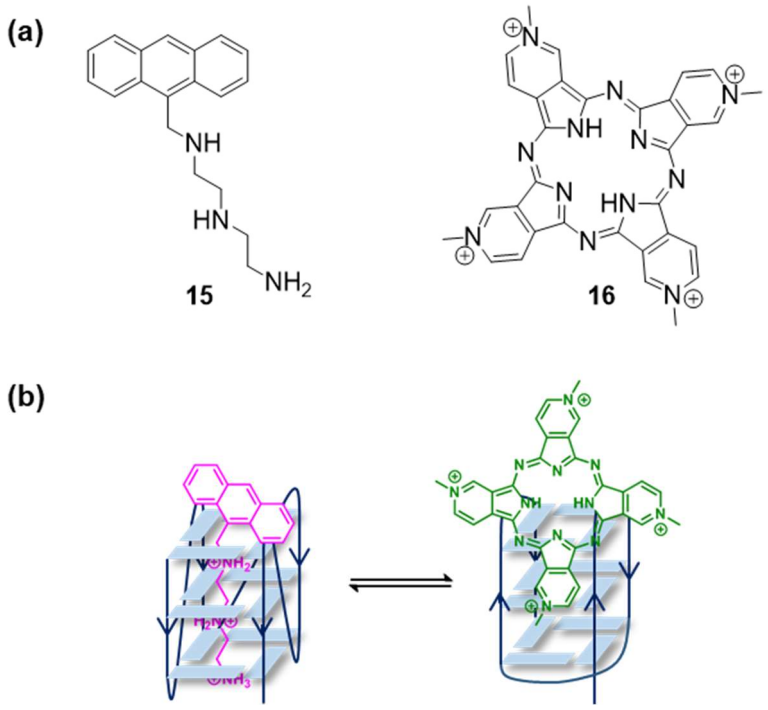


Figure 1.18: Reversible ligand-mediated control of G4 structure. (a) chemical structures of triaminoanthracene **15** and porphyrazine **16**; (b) toggling of G4 between parallel and antiparallel G4 topologies of telo24-Na^+ by sequential addition of the ligands. *Figure (b) adapted from the author's publication with permission © Wiley*.[1]

1.3.2.3 Ligand triggers

The vast majority of G4 ligand development has been undertaken with a therapeutic goal in mind, as discussed previously in Section 1.3.1. During these studies, ligands are often screened for binding to pre-formed G4 structures formed under ion-rich conditions, under which a major folded nucleic acid conformation predominates. However, especially in the case of the highly polymorphic telomeric DNA, certain ligands are actually able to induce changes in the G4 folding topology from the native conformation to a different folded structure, providing an additional opportunity for the control of G4 architectures.[1] As well as inducing the formation of G4 from an unfolded structure in the absence of stabilising metal cations,[143] some small

G4-targeting molecules exert folding changes in the presence of metal ions, thereby displaying potential to control G4-based topology in physiologically-relevant conditions where, unlike the applications discussed in the previous section, the concentration of metal ions cannot be readily controlled.

One of the first examples of a ligand-driven G4 nanodevice was reported by Balasubramanian and co-workers in 2007.[144] A triamino-anthracene **15** (Figure 1.18a) derivative was designed to bind G4 by simultaneous π-π stacking interactions with the surface G-tetrads and threading of the side-chain through the central channel of the G4. The spacing of the amine groups was chosen to mimic the distance between K^+ ions in the native G4, allowing the amine moieties to act as the channel cations when protonated at physiological pH. Interestingly, this ligand induces a conformational switch of the telo24 G4 from an antiparallel structure in sodium buffer to a parallel form. Furthermore, since the porphyrazine ligand **16** (Figure 1.18a) preferentially stabilises the antiparallel form of telo24, the G4 conformation could be toggled between antiparallel and parallel forms by sequential addition of the two ligands (Figure 1.18b). However, the conformational switch was only reversed once in this manner; possibly further cycles were inhibited by an overaccumulation of the ligand fuel, and no specific application has yet been developed at the time of writing.

More recently the Zhou group employed a host/guest approach towards the ligand triggered control of G4 structure towards regulation of the enzymatic activity of thrombin.[145] This system is based on the affinity of the terminal piperidinium moieties of azobenzene derivative **19** for a cucurbit[7]uril (CB7, **17**) host (Figure 1.19a). Whilst free ligand **19** induces the formation of a G4 structure in a G-rich telomeric strand, sequestration by CB7 **17** causes dissociation of the ligand from the G4 and concomitant unfolding of the DNA secondary structure occurs. By fuelling the system alternately with CB7 **17** and a competitive guest, (adamantylammonium **18**) 13 switches of G4 unfolding and refolding were achieved. To regulate enzymatic activity, an aptamer containing a G4-forming linker (based on telomeric TTAGGG repeats) that binds to and inhibits thrombin activity was employed (Figure 1.19b). The addition of ligand **19** triggers the folding of the linker thereby causing the dissociation of the inhibitor, un-gating the activity of the thrombin. The enzyme can be subsequently deactivated by adding CB7 **17** which sequesters the ligand, causing unfolding of the G4 and regeneration of the thrombin/inhibitor complex. In the same way as the original proof-of-concept outlined above, addition of adamantylammonium cation **18** reverses this process.

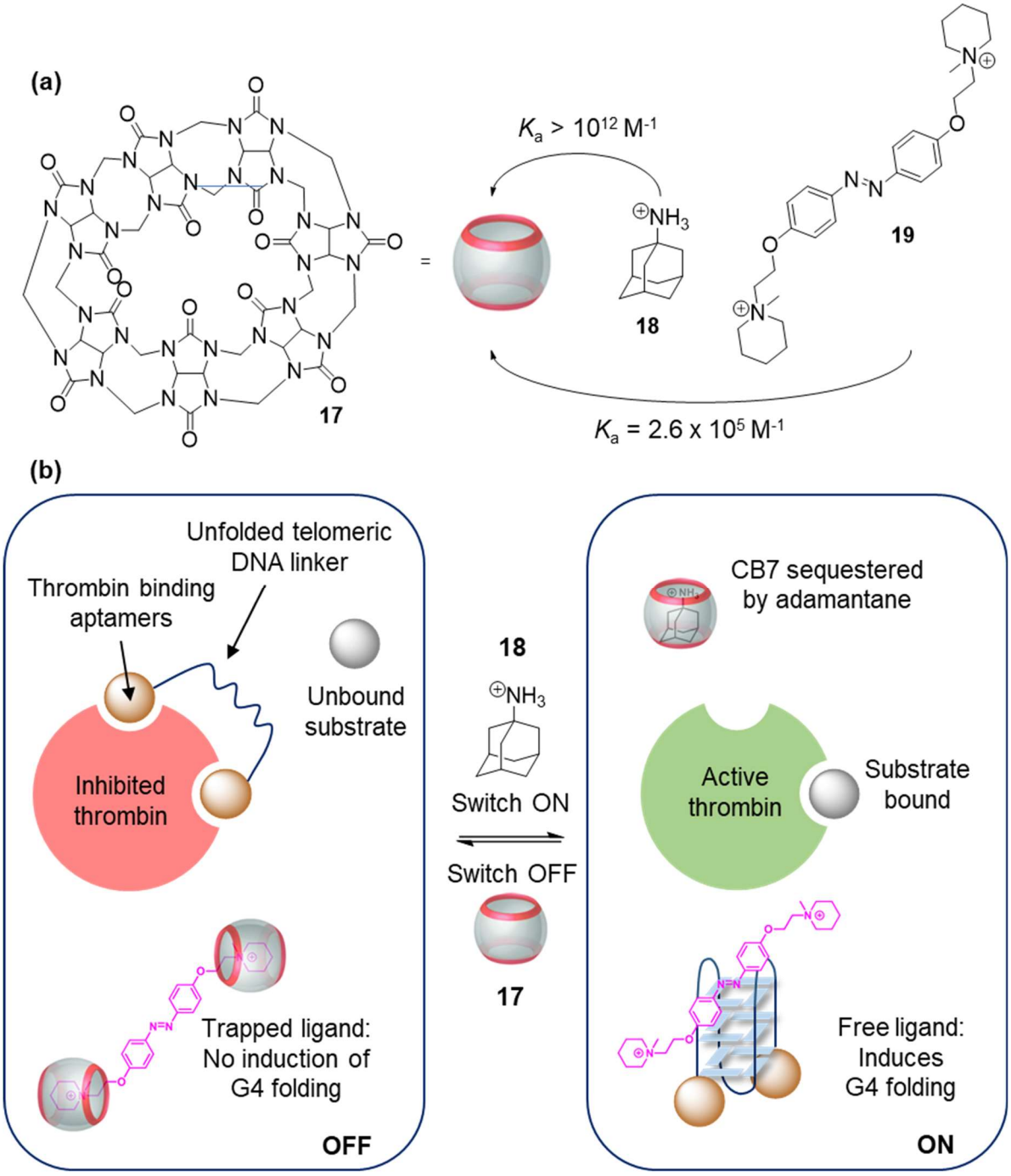


Figure 1.19: Ligand-triggered regulation G4 structure to control the activity of an enzyme using a host/guest approach. (a) Chemical structure and cartoon representation of cucurbit[7]uril **17**, an efficient molecular host for ammonium cations such as **19** and adamantylammonium **18**. (b) Switchable control of thrombin activity by exploiting the host/guest chemistry of G4 ligand **19** and CB7 **17**. *Figure adapted from the author's publication with permission © Wiley.*[1]

1.4 Photo-control of G4 DNA

The previous sections introduce the two main opportunities afforded by regulation of G-quadruplex DNA: regulation of formation *in vivo* as a potential therapeutic strategy, or regulation of the structure in order to generate a mechanical or spectroscopic output as the basis of a functional system. In all of the systems discussed so far, the control over G4 is achieved using exclusively chemical triggers. Though this is not a problem in some cases, the reliance on chemical stimuli alone may lead to undesirable effects, impracticalities or limitations in real-word applications. In therapy, drugs are often administered intravenously and are therefore distributed throughout the whole organism, rather than just to the desired site of action, and this is partially responsible for side-effects due to off-target activity in other vital organs. In the case of functional applications, such as the development of sensors or other responsive materials, the sequential addition of chemical fuels must be precisely balanced in order to allow robust control over the system and, depending on the application, may not always be convenient or possible. Meanwhile, light offers several complementary advantages over chemical stimuli for the regulation of both biological and functional systems. In this section, the potential to regulate G-quadruplex structures using light towards therapeutic and functional applications is considered.

1.4.1 Light as a stimulus for the regulation of biomolecule structure and function

Upon the irradiation of an organic chromophore with light at appropriate wavelength, an electron is excited from the highest-occupied molecular orbital to the lowest-unoccupied molecular orbital. This transition allows various types of chemistry that are either not possible in the ground state, either on orbital symmetry grounds or because the thermal barrier to interconversion is too high. For example, the rotation about a stilbene double bond does not occur thermally at ambient temperature, meaning the *E* and *Z* isomers are stable and do not interconvert between one another in the ground state, but this barrier is easily surmounted in the photoexcited state, providing a pathway for thermal relaxation to the other isomer (Figure 1.20).[146]

Light is an attractive means for the regulation of both biological and functional systems since it can be administered with much higher spatiotemporal precision than chemical fuels.[147–151] By controlling the wavelength, intensity and irradiation time, both the energy and dose of photons can be controlled. Unlike chemical stimuli for which addition is, in practice, normally irreversible, irradiation does not contaminate the system (assuming no unwanted side-reactions take place) and therefore is traceless once irradiation has ceased.

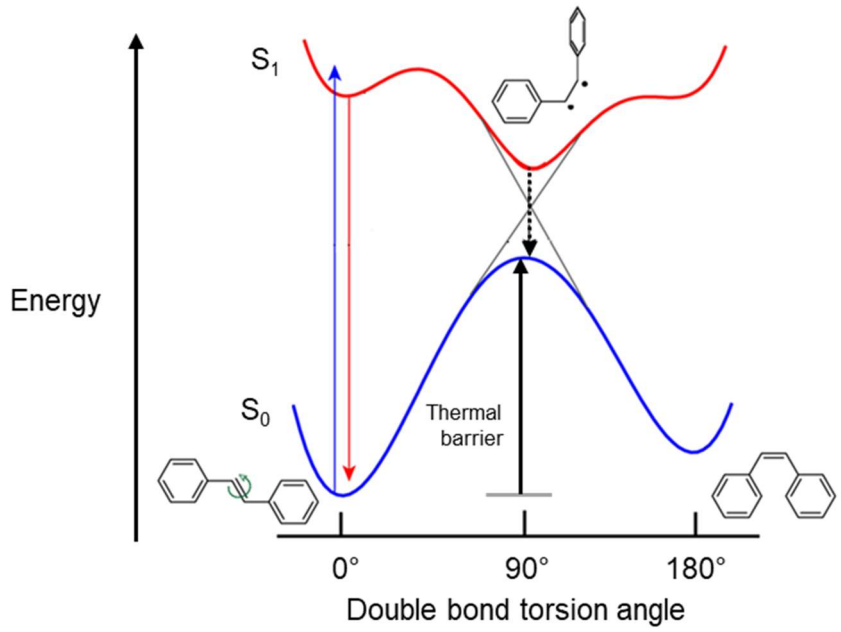


Figure 1.20: Example of a molecular photoswitch – the isomerisation of *E* to *Z* stilbene. In the ground state (S_0) the thermal barrier to thermal interconversion (blue trace) is too high for the transformation to occur at ambient temperature. Photoexcitation to the higher energy (S_1) state allows this energy barrier to be surmounted.

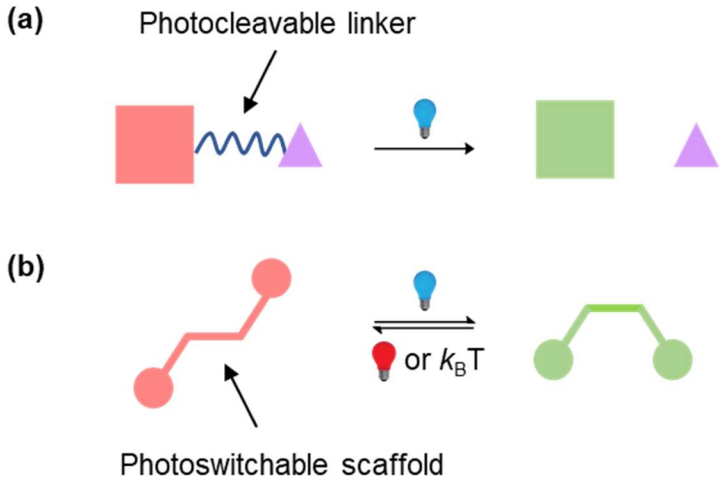


Figure 1.21: Approaches to the photocontrol of biomolecules. (a) the activity of the molecule is blocked (red) by conjugation to caging group by a photolabile linker. Photocleavage releases the active (green) form of the molecule; (b) the structure of the molecule is switched between a comparatively inactive (red) and active (green) form by light; in some cases, the process is thermally reversible.

A variety of photoresponsive chemotypes have found application in different biological settings.[151] These fall into two main categories: photocages and photoswitches (Figure 1.21).[147] In a photocaging strategy, the biological activity of the molecule is deactivated by conjugation by a photolabile group to a fragment which inhibits its activity. Photoirradiation cleaves the caging group, delivering the molecule in the active form (Figure 1.21a). In the case of photoswitches, irradiation with light causes a change in the molecular structure. In many cases this can be reversed either thermally or with light of a different wavelength (Figure 1.21b). This approach is particularly attractive since no waste products are generated and (assuming the two photoisomers have different activities in the system of interest) the activity of a molecule may therefore be switched reversibly without the need for sequential addition of chemical fuels, offering additional levels of control over the system. The deployment of molecular photoswitches in the regulation of biological systems has been extensively reviewed by others, and four primary scaffolds emerge as the most common: azobenzenes, stilbenes, diarylethenes and spiropyrans.[151] The general structures and photoisomerisation reactions of these chemotypes are shown in Figure 1.22. The azobenzene (*E*-**20**/*Z*-**20**) and stilbene (*E*-**21**/*Z*-**21**) chromophores undergo *E/Z* isomerisation, whilst the diarylethene (**22o/22c**) and spiropyran (**23o/23c**) scaffolds are converted between closed and open forms.

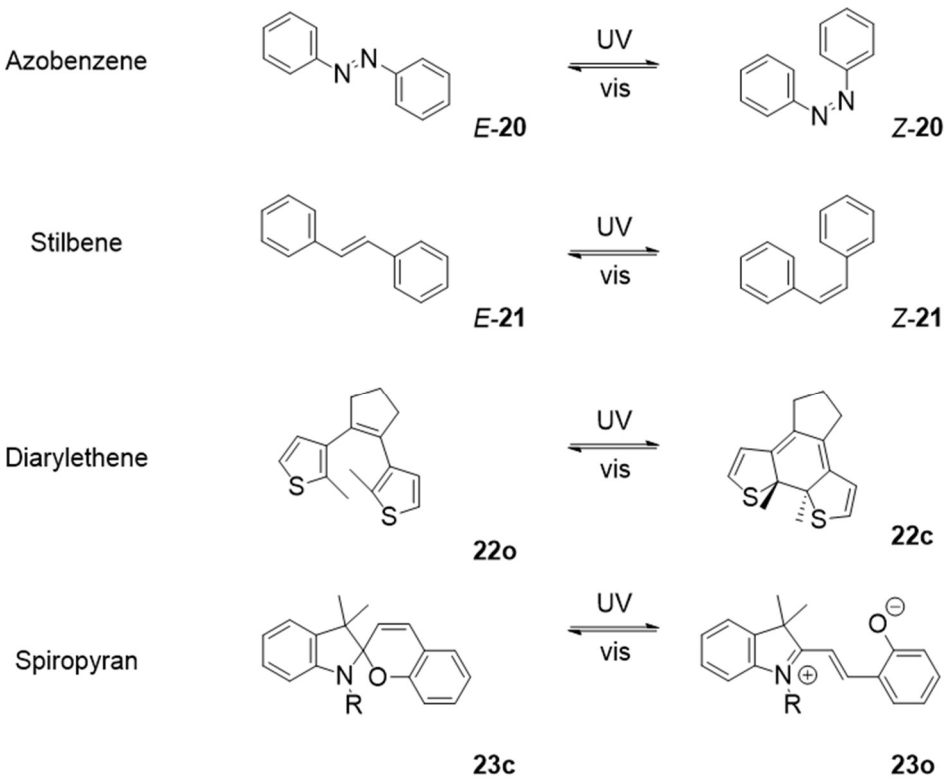


Figure 1.22: Organic photoswitches most commonly deployed in biological applications.

Regarding the application of photochemistry towards the regulation of nucleic acid structures, two possible approaches can be considered. Either the photoresponsive functionality may be pre-incorporated in the biomolecule itself, or a ligand can be designed where binding activity can be controlled either by a photocaging or photoswitching strategy. The former example is particularly useful in the development of functional materials which need not rely on natural DNA structures and can be easily modified to include the necessary functionality. The latter is more relevant in biological contexts, such as therapy, where pre-modification of the target is not always possible, and control is instead exerted using small molecule ligands. These two strategies are considered below in the context of the opportunities afforded by the control of G-quadruplex structure.

1.4.2 Photocontrol through incorporation of photoswitches

To date, the majority of studies concerned with the photoregulation of nucleic acid structures focus the control of duplex DNA hairpins.[149,152] Meanwhile, examples of photoresponsive control of G4 structures in a similar way are rather rare despite the proven applications of this alternative architecture in the development of functional materials. Ogasawara and Maeda demonstrated the photoresponsive formation and unfolding of G-quadruplex by installation of a modified guanine nucleobase decorated with the photoisomerisable 8- fluorenylvinyl group (*E*-**24**/*Z*-**24**, Figure 1.23).[153] In the *E* state the G-quadruplex structure assembles, but isomerisation to the *Z* state using 410 nm light causes the unfolding of the structure. The folding and refolding process could be cycled several times, and in turn control the affinity of the DNA structure for thrombin. Thus, the system achieves a similar end to the hybridisation-controlled catch-and-release of thrombin reported by Simmel (see Section 1.3.2.1),[139] but is controlled simply by photoirradiation rather than requiring sequential addition of oligonucleotide fuels. Later, the Heckel group demonstrated the photoresponsive formation of an intermolecular G-quadruplex by incorporation of azobenzene functionality into the backbone of guanosine tetramers (*E*-**25**/*Z*-**25**, Figure, 1.24).[154] Very recently, the Tucker group engineered a complementary system where thrombin binding to G-quadruplex is modulated by photodimerisation of two anthracene moieties tethered to the oligonucleotide chain (Figure 1.25).[155]

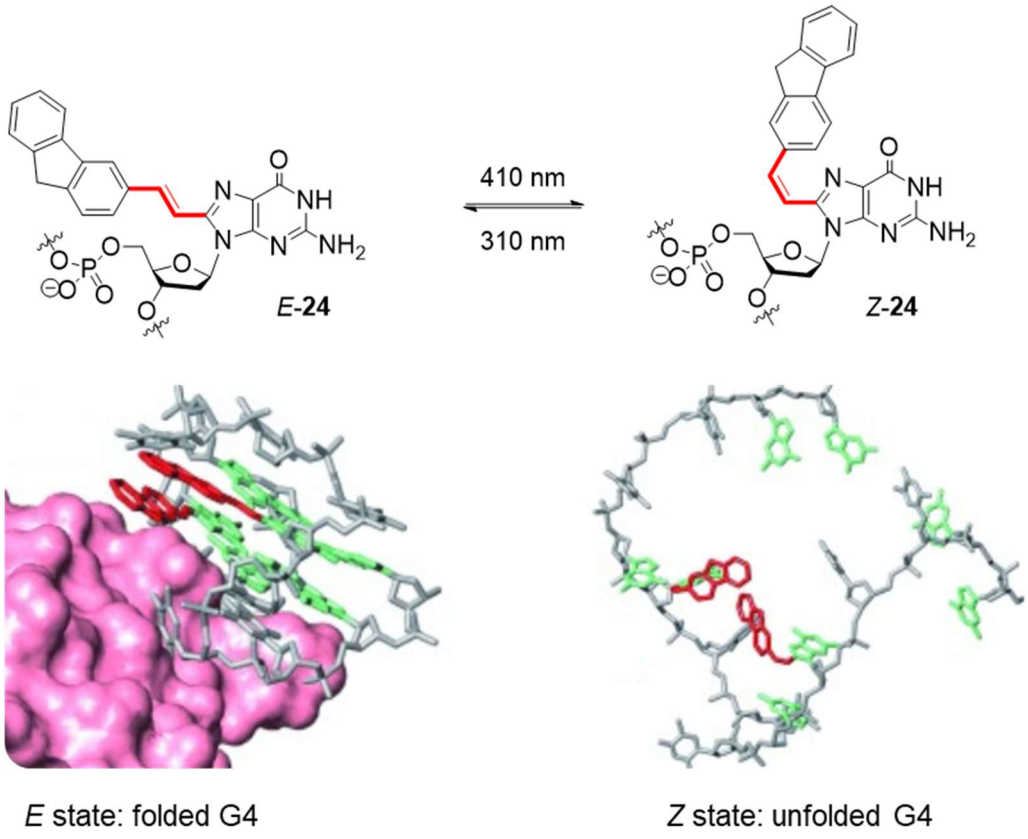


Figure 1.23: Photoswitching of G-quadruplex folding by incorporation of the photoresponsive 8-fluorenylvinyl group, developed by Ogasawara and Maeda. *Structural figures reproduced with permission © Wiley*.[153]

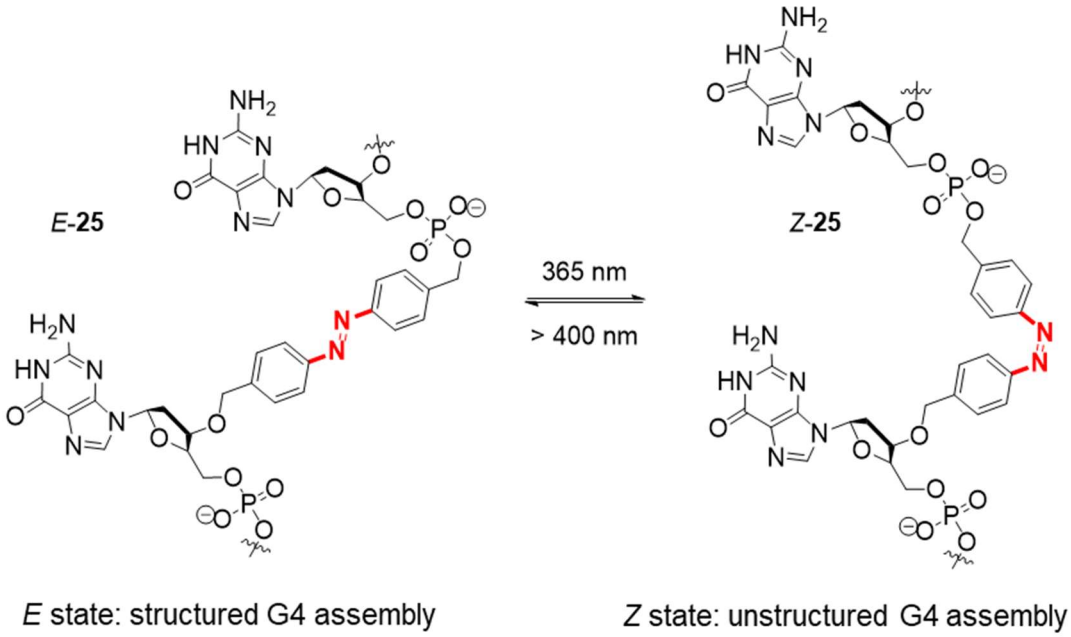


Figure 1.24: Photoresponsive formation of an intermolecular G-quadruplex from azobenzene-linked nucleotides, developed by Heckel and co-workers.

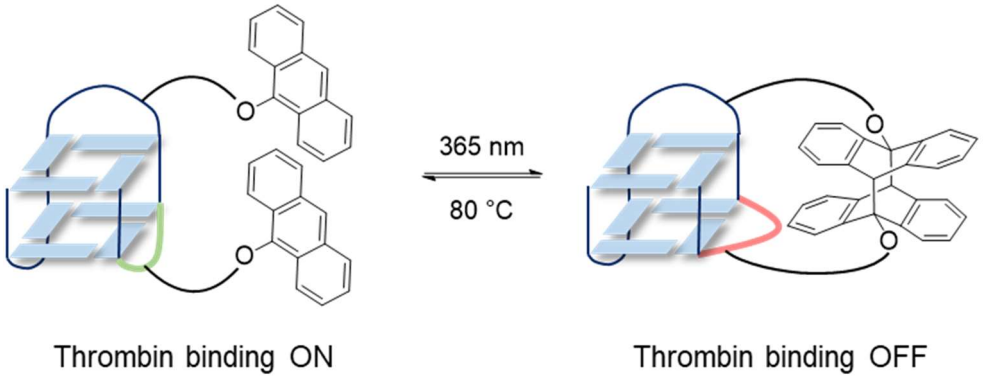


Figure 1.25: Control of G4/thrombin binding by anthracene photodimerisation, developed by Tucker and colleagues.

Though many exciting applications of these elegant approaches to regulating G4 structure can be envisaged, the drawback in these cases is the need to rely on pre-incorporation of unnatural functionality into the oligonucleotide, limiting their potential applications in biological setting. Complementary applications, where such a modification of the underlying DNA sequence is undesirable, could be achieved by using a supramolecular approach, through the design of photoresponsive small-molecule ligands.

1.4.3 Photocontrol using photoresponsive ligands

Regarding the development of photoresponsive small molecules for use in biological applications, much interest has been devoted to the development of photoswitchable inhibitors of proteins, towards the goal of achieving light-responsive therapeutic agents.[148,156–158] This approach, known as photopharmacology, aims to reduce the off-target effects resulting from distribution of the drug around the whole organism by using light to activate the molecule specifically at the site of action, and subsequently deactivating the compound by light and a different wavelength, or a thermal process (Figure 1.26). This approach requires a photoswitchable ligand where there is a significant difference in bioactivity between the two isomers. For example, Feringa and colleagues developed a photoresponsive inhibitor of a histone deacetylase based on an azobenzene core, which was 39-fold more active in the *Z* versus the *E* state (*E*-**26**/*Z*-**26,** Figure 1.27) which led to increased toxicity to HeLa cells.[159]

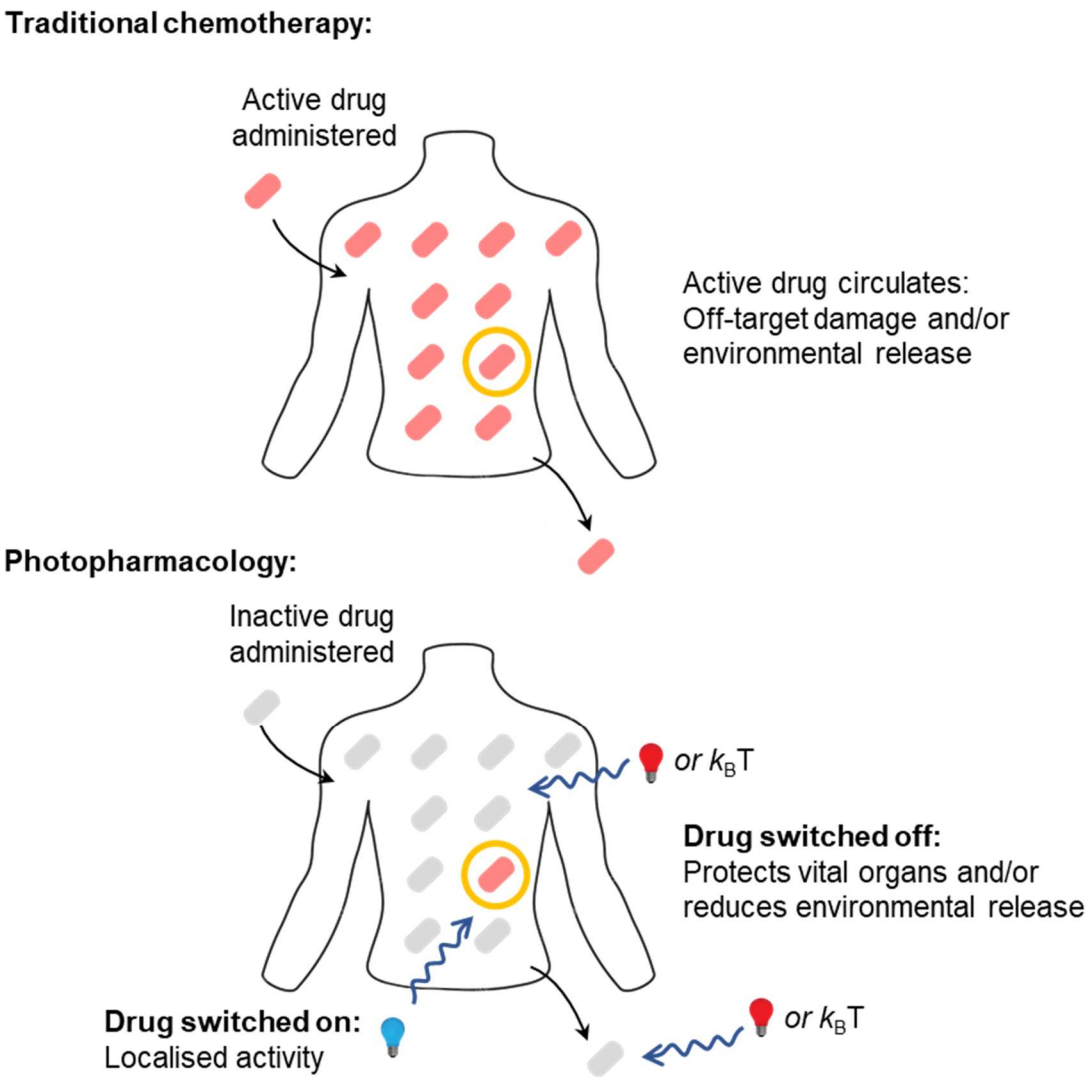


Figure 1.26: Concept of photopharmacology. (a) in traditional chemotherapy; (b) *in situ* control of activity through use of photoswitchable drugs.

Despite DNA being well-studied as a therapeutic target, photopharmacological approaches towards targeting DNA have barely been explored. The only examples to date consider duplex DNA binding. For example a photochromic spiropyran (**27o/27c**, Figure 1.28) developed by Andreasson and co-workers was shown to bind calf thymus DNA in the planar open (**27o**) zwitterionic state, whilst binding was deactivated upon photocyclisation to the closed (**27c**) state.[160] In a subsequent study, the authors evaluated the uptake and cytotoxicity of the two isomers in HEK 293 cells, observing that *in cellulo* photoisomerisation of the compound led to significantly increased (52% versus 3%) cell death.[161]

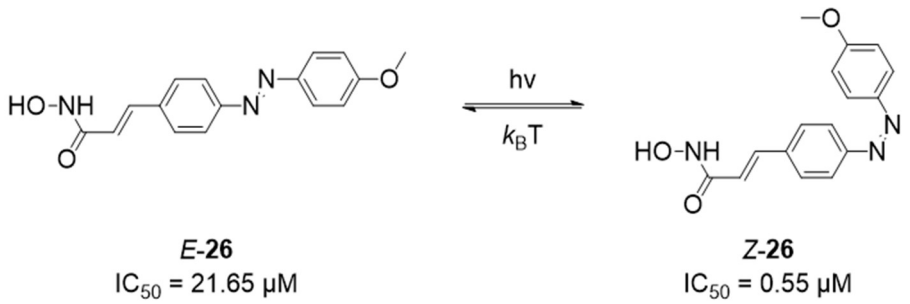


Figure 1.27: A photoswitchable histone deacetylase inhibitor developed by Feringa and co-workers.

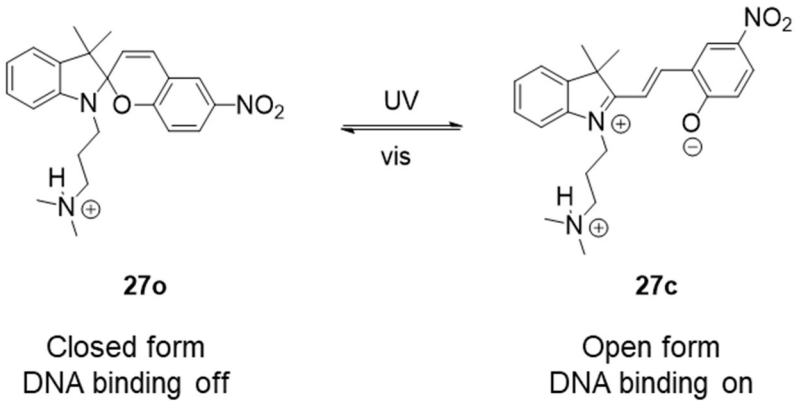
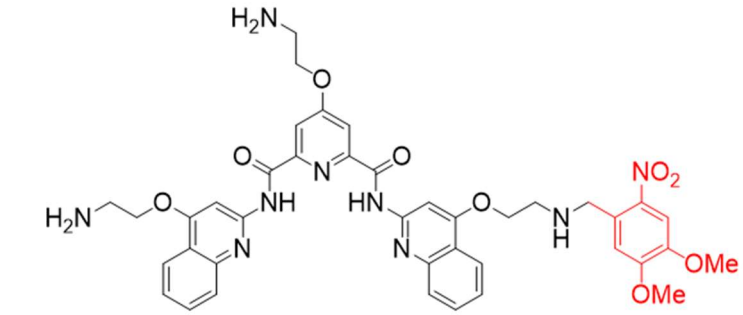
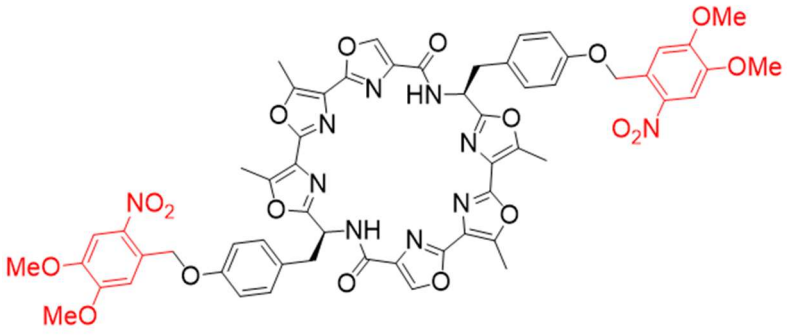


Figure 1.28: A spiropyran-derived photoresponsive ligand for duplex DNA, developed by Andreasson and colleagues.

Though quadruplex DNA has long been studied as a therapeutic target, and significant therapeutic leads have been generated which demonstrate the benefit of targeting these structures selectively over duplex DNA (Section 1.3), to date no attempts have yet been made to explore the potential of these non-canonical structures as targets for photopharmacology. Two examples of photocaged G-quadruplex ligands have been reported, one based on pyridostatin (**28**)[162] and the other based on a telomestatin mimic (**29**, Figure 1.29).[163] In both cases, the G-quadruplex binding activity of the active compound is neutered by the presence of the nitroveratryl group which may be removed *in cellulo* by irradiation with UV-light, resulting in the unmasking of the compounds' toxicity. However, the photocaging approach permits only one-way control of the ligand activity.



28



29

Figure 1.29: Photocaged G-quadruplex ligands. The nitroveratryl caging group (shown in red) is removed by irradiation with 385 nm light.

The only example of reversible control of G4 with a photoswitchable small molecules is azobenzene derivative **19** (previously introduced in Section 1.3.2) first reported by Wang and co-workers in 2010 (Figure 1.30).[143] The compound was found to induce G4-folding in the absence of cations in water in the *E* form, whilst the *Z* isomer did not have such an effect. The unfolding and refolding of G4 could be cycled at least ten times without any appreciable photo-fatigue by photoisomerisation of the ligand (Figures 1.30b and 1.30c). Subsequently, the authors applied the photoswitchable ligand to regulate the activity of thrombin in a manner analogous to the host-guest approach discussed in Section 1.3.2 (Figure 1.30d).[164] Meanwhile, the authors attempted to exploit this system in the presence of metal ions to mimic physiological conditions, presumably with biological applications in mind.[165] Unfortunately, compound **19** was found to be ineffective in these studies. The authors do not specifically comment on why this system is adversely affected by the presence of metal ions, though perhaps the stabilisation of G4 folding topology induced by the metal cation is more difficult to overcome by ligand-driven effects.

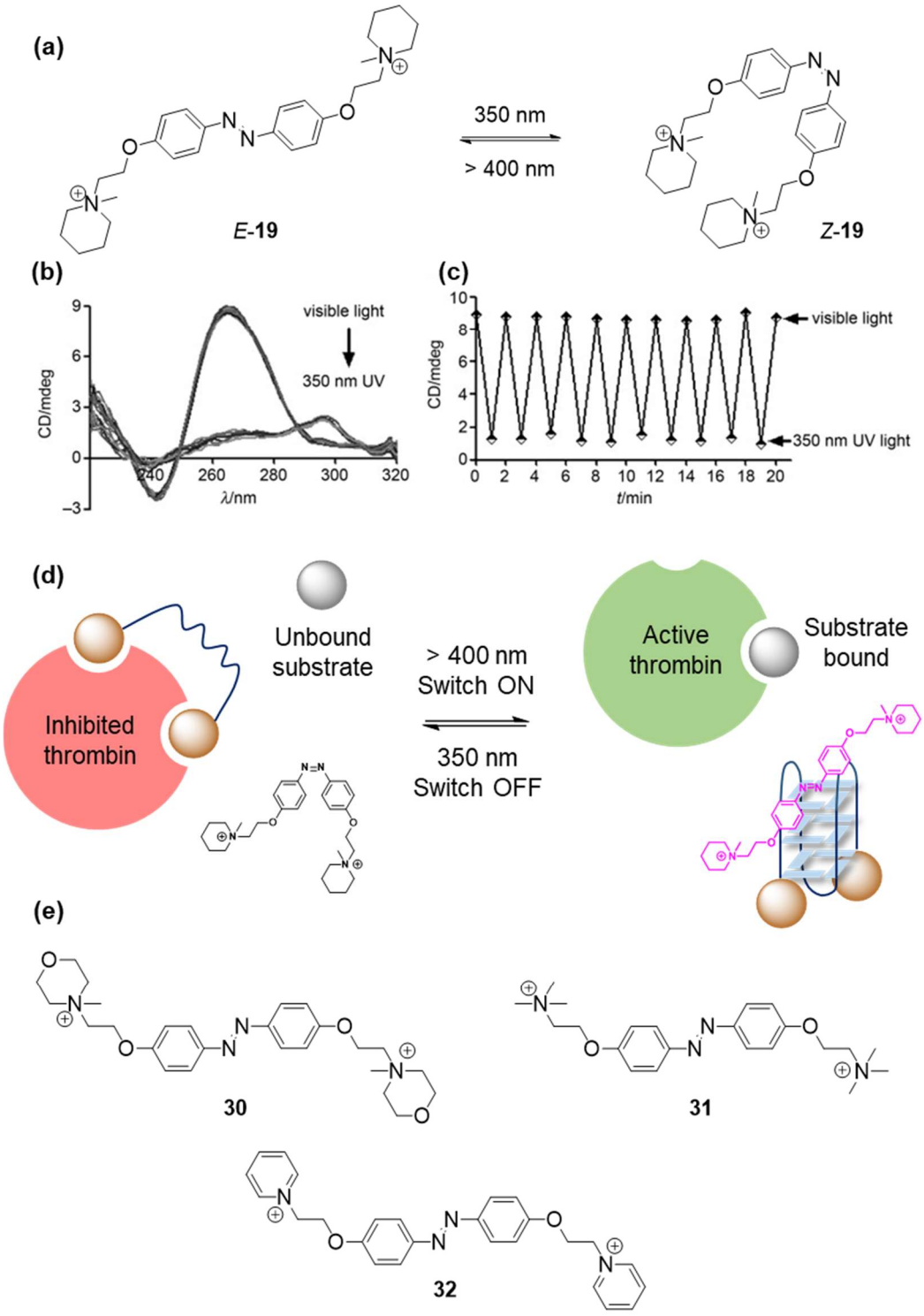
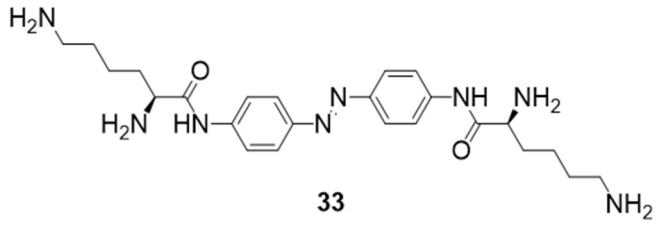
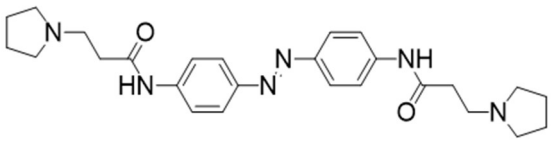


Figure 1.30: Photoswitchable control of G4 folding by azobenzene derivative **19**. (a) ligand photoisomerisation between *E* and *Z* forms; (b)-(c) reversible G4 folding/unfolding triggered by ligand photoisomerisation monitored by circular dichroism spectroscopy; (d) application to photoregulation of thrombin activity; (e) second-generation analogues aiming to photocontrol G4 structure in the presence of metal ions. *Figure (b) used with permission © Wiley*.[143]

This led the authors to develop further analogues of **19** (**30**-**32**, Figure 1.30) which appeared to allow the photo-regulation of G4 topology in potassium-rich conditions to some extent. Based on apparent changes in the characteristic circular dichroism spectra, the authors proposed the *E*-forms of the ligands induced a switch in topology from hybrid to antiparallel in the presence of potassium ions. However, details about the nature of the ligand-induced structures or changes in binding mode are unclear in the absence of further structural information.



33



34

Figure 1.31: Further azobenzene derivatives investigated as photoresponsive G-quadruplex ligands.

Further azobenzene-derivatives functionalised with amino acid side chains were recently reported by the Matczyszyn group, such as lysine derivative **33** (Figure 1.31).[166] Though in these cases the formation of the G4/ligand complex could be controlled by the photoisomerisation process, the compounds appear to exert little effect on G4 topology and also appear to interact significantly with duplex DNA, perhaps indicating that applications of azobenzene derivates may be confined to situations where a high level of quadruplex/duplex selectivity is not required. A series of closely-related azobenzene-derived ligands conferred only very modest thermal stabilisation to G4 structures: even at high ligand concentrations (40 μM) only 4 °C thermal stabilisation was induced in G4 for the most potent analogue reported (**34**, Figure 1.31) in potassium-rich conditions, indicating that this chemotype does not display high affinity for G4.[167] Meanwhile, azobenzene derivatives are demonstrated to intercalate and stabilise duplex DNA at comparable concentrations,[168] suggesting that this chromophore is more likely suited to intercalating DNA base pairs rather than targeting the specific features of

G4. Furthermore, at the time of writing, the potential of these compounds to target G4 for therapeutic effect has not yet been investigated.

1.5 Conclusions and thesis aims

The discussions in the previous sections highlight the myriad potential applications afforded by the control of G4, both in therapy and in the development of nanodevices. Though progress has been made towards both of these areas, opportunities for further development remain. In particular, photoresponsive G4 ligands would offer interesting means to control G4 topology in the development of molecular devices or as the basis of G4-targeted photopharmacology strategies. Though some progress has been made towards regulating G4 architectures with photoswitchable small molecules, this research area remains significantly underexplored. In particular, issues such as G-quadruplex/duplex selectivity and potency in physiological conditions require further investigation and development. Azobenzenes are the only photoswitches to date investigated for binding to G4 DNA, despite the availability of several other potential photoresponsive scaffolds which may furnish ligands that provide complementary effects.

The overarching goal of this research project was therefore to explore the development of novel photoresponsive G4 ligands that function effectively in physiologically-relevant conditions and explore their potential as tools for G4-targeted photopharmacology or controlling G4-based nanodevices. The results of these endeavours are reported in the following chapters:

Chapter 2 describes the design, synthesis, biophysical and cellular evaluation of a new class of amine-decorated G4 ligands derived from stiff-stilbene, the first example of DNA targeting molecules derived from this particular molecular photoswitch. This work acts as an initial proof of concept that this type of scaffold can be used to target G4 with good selectivity in biologically-relevant conditions. The chapter also serves to introduce the various techniques used to study G4/ligand interactions.

In Chapter 3, the design of the stiff-stilbene ligands is developed to improve the ligand potency. The synthesis and evaluation of *N*-methylpyridinium derivatives is reported, which demonstrate improved G4 recognition properties compared to the first-generation amine-decorated analogues. This work identifies the ligands as tools to perturb G4-folding in physiologically-relevant conditions and that the stilbene configuration is a critical determinant of G4 ligand activity. The ligands are screened for activity against a small panel of disease models, identifying correlation between G4 binding activity and toxicity.

In Chapter 4, the photochemistry of *N*-methylpyridinium G4 ligands is explored towards the goal of exploiting the difference in activity of the photoisomers in photopharmacology or control of G4 architectures. Whilst the ligand photochemistry in water was ultimately unsuited towards photopharmacological applications, the reversible folding and unfolding of a G4 structure is demonstrated in sodium-rich conditions.

Chapter 5 reports on the design, synthesis, biophysical screening and cellular evaluation of G4 ligands designed from the dithienylethene chromophore. These ligands are found to undergo robust photoswitching between two photoisomeric states in buffered conditions, which in turn exhibit different binding modes to G-quadruplex DNA and toxicity to a cancer cell model.

Chapter 6 summarises the results of the thesis, offering general conclusions and suggestions for future work.

Chapter 7 contains details of experimental procedures, including compound synthesis and characterisation.

A removable glossary of frequently-discussed ligands is included to assist the reader. This is also printed at the back of the thesis.

2 Amine-decorated stiff-stilbene G4 ligands

This chapter is based upon work published in *Chemistry: A European Journal* of which I am first author.[169] I was responsible for all elements of ligand design, synthesis and biophysical studies, which were undertaken at the University of Bristol, UK. Compounds A and C were synthesised by Rosina S.L. Fisher, an MSc student working under my direct supervision. I performed the *in vitro* assays at the Institute of Parasitology and Biomedicine (CSIC), Granada, Spain, with assistance from Dr Pablo Peñalver.

2.1 Introduction

This chapter describes initial progress towards the development of a novel series of G4 ligands derived from a photoresponsive stiff-stilbene chromophore. The rationale for the choice of ligand scaffold is first explained in the context of previously reported small-molecule photoswitches, general G4 ligand design principles and the few reports of previous attempts to engineer photoresponsive G4 ligands. Subsequently, the synthesis of a library of compounds is reported, designed to test the general viability of the scaffold to serve as the basis of new G4 binding agents, and to evaluate the effect of ligand isomerisation on G4-binding activity. The initial screening of the compound library for binding to G4 structures using a thermal melting assay is then described. These results provide preliminary indications of structure-activity relationships present in the new ligand series and identify a proof-of-concept ligand pair that indicates G4-binding activity is determined by the isomeric configuration of the central stilbene moiety. This promising ligand pair is then investigated in greater detail, using a combination of circular dichroism, UV/visible absorbance and NMR spectroscopies, to validate this preliminary result and gain further insight into the nature of G4 recognition by the novel ligand scaffold. Finally, the potential to exploit the difference in activity of the ligands towards therapeutic ends is explored by screening the compounds for anticancer and antiparasitic activity in *in vitro* assays.

2.2 Ligand design and synthesis

2.2.1 Consideration of ligand scaffolds

The first step in the design of novel photoresponsive G4 binding agents was the selection of a photoresponsive chromophore that could be readily adapted to serve as the basis of a new series of high-affinity G4 ligands. Considering the photochromic scaffolds previously introduced in the context of general G4 ligand design principles (Section 1.3.1.4) the most promising scaffolds to explore as the basis of photoresponsive G4 ligands are the azobenzene

(*E*-**20**/*Z*-**20**) and stilbene (*E*-**21**/*Z*-**21**) photoswitches (Figure 1.22). These scaffolds undergo geometrical change upon isomerisation, adopting a planar arrangement in the *E* form whilst being non-planar in the *Z* form owing to steric repulsion between the *ortho* aromatic hydrogen atoms.[170] Chemotypes derived from planar aromatic scaffolds are generally privileged in G4 ligand development since, in many cases, they are found to effectively bind G4 through π-stacking interactions with the planar G-tetrads.[44] The change in geometry from planar to non-planar forms triggered by photoisomerisation can therefore be expected to significantly alter the ability of the ligand to form such interactions. Other potential G4 ligand binding modes, such as loop and groove binding, are also likely to be affected by such changes in spatial arrangement of the ligand atoms. For these reasons, investigation of the G4 binding properties of diarylethene (**22o/22c,** Figure 1.22) derivatives was not initially pursued, since the molecule undergoes comparatively little structural change upon photoisomerisation.[171] However, the diarylethene scaffold was studied at a later point in the project and delivered some promising results (Chapter 5). Spiropyrans (**23o/23c,** Figure 1.22) also have potential as photoresponsive G-quadruplex binding molecules, based on the difference in shape between the closed form (containing a tetrahedral carbon atom at the spirocentre) and open form (which adopts a more planar structure). This scaffold was not explored in the present project but could be investigated in future work.

2.2.2 From azobenzene to stiff-stilbene

As described previously (Section 1.4.3), work has previously been undertaken to deploy azobenzene derivatives as photoresponsive ligands to facilitate the responsive control of G4 structure. Though promising results have been obtained,[143,164] this approach suffers from several limitations that likely prevent the use of this system for *in vivo* applications (e.g. photopharmacology) of G4 targeting. Though the switching of G4 conformation functions effectively in the absence of metal ions, the compound is not effective in influencing G4 topology cation-rich conditions that mimic the intracellular environment.[165] Furthermore, a series of closely-related azobenzene-derived ligands conferred only modest thermal stabilisation to G4 structures: even at high ligand concentrations (40 μM) only 4 °C thermal stabilisation was induced in G4 for the most potent analogue in potassium-rich conditions, indicating that this chemotype does not display high affinity for G4 under physiologically-relevant conditions of high ionic strength. These issues may explain why compound **19** (in the active G4-binding *E* configuration) does not demonstrate toxicity (EC_{50} > 125 μM) to HeLa cells.[164] Meanwhile, azobenzene derivatives are demonstrated to intercalate and stabilise duplex DNA at comparable concentrations, suggesting that this chromophore is more likely suited to intercalating DNA base pairs rather than targeting the specific structural features of

G4.[168] Considering the azobenzene core in the context of the best-in-class G4 ligands previously reported (e.g. PhenDC3 **10**, Figure 1.14), it was reasoned that the poor G4 affinity of azobenzenes is a result of comparatively weak stacking interactions between the ligand π system and the G-tetrads, the binding mode suggested by computational investigations in the original study.[143] Whilst the *E* azobenzene can adopt a planar conformation, it possesses only two aromatic rings and hence a relatively small hydrophobic surface area. Furthermore, rotation about the *N*-aryl bonds permits conformational freedom in solution, resulting in an entropic penalty to adopt the active conformation to bind to the macromolecule. Therefore, stacking interactions between the azobenzene ligand and the G-tetrads can be expected to be weaker than those present in the more potent G4 ligands reported to date, many of which are derived from fused ring systems that provide greater hydrophobic stacking surface and increased rigidity.

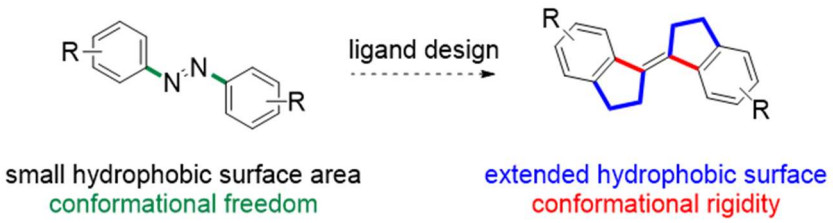
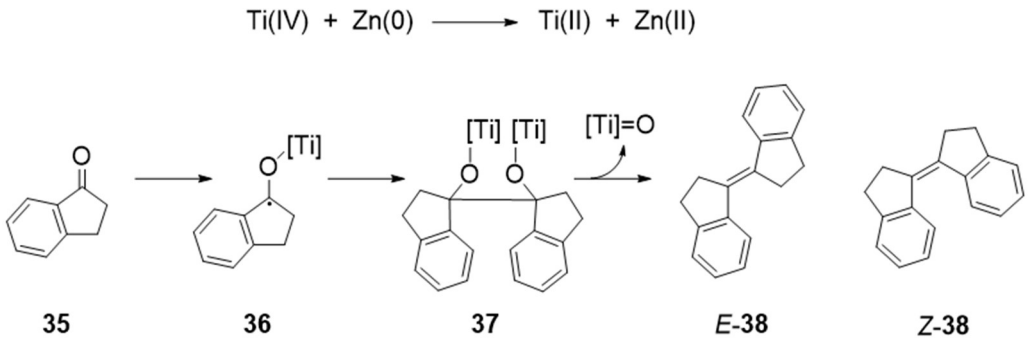


Figure 2.1: Design concept: from azobenzenes to stiff-stilbenes.

In light of the above considerations, it was reasoned that G4 binding could be markedly improved by extending the hydrophobic stacking surface and increasing the conformational rigidity of the ligand scaffold, whilst maintaining the feature of photoresponsive functionality. In pursuit of this goal, stilbenes presented an interesting opportunity, as the possibility of tetrasubstitution of the central double bond in these variants allows the generation of bicyclic systems that would feature both increased hydrophobic surface area and the additional rigidity that were anticipated to be beneficial for more effective recognition of G4 (Figure 2.1).

The synthesis of such conformationally restricted stilbenes, colloquially termed stiff-stilbenes, has been previously reported.[172] Commonly, the McMurry coupling is employed (Scheme 2.1).[173] In this reaction, titanium(IV) is reduced *in situ* by zinc to generate a low-valent Ti(II) species, which acts as the reducing agent of a carbonyl compound. In one proposed mechanism, single electron reduction of the carbonyl group of compound **35** generates a ketyl radical **36**, which subsequently dimerises to form pinacol-type product **37**.[174] Elimination of the oxo-titanium species furnishes the desired products (*E*-**38** and *Z*-**38**), normally as a mixture of stereoisomers, with the *E*-isomer generally predominating on steric grounds.[173]



Scheme 2.1: Proposed mechanism of stiff-stilbene scaffold formation by McMurry coupling.

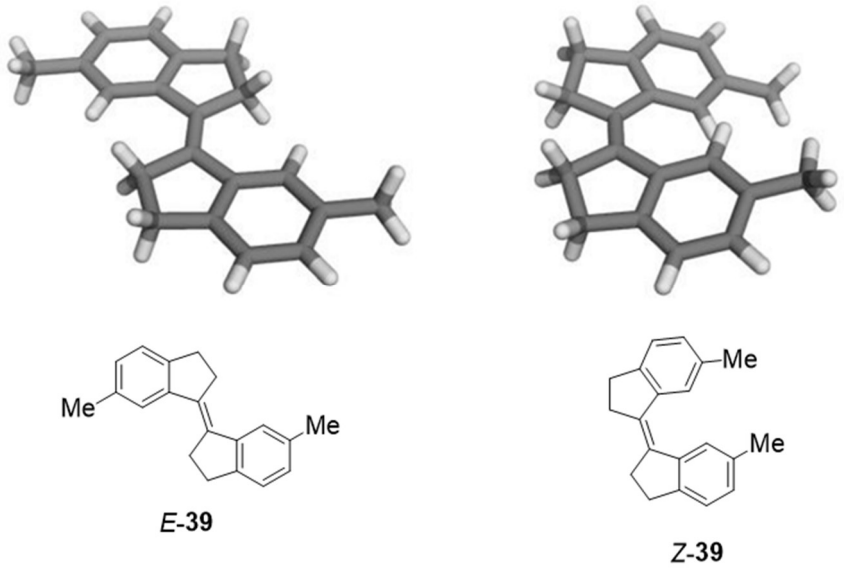


Figure 2.2: X-ray crystal structures of substituted stiff-stilbenes *E*-**39** and *Z*-**39**. *Structural figures reproduced with permission*.[172]

The photophysical and photochemical behaviour of a range unfunctionalised stiff-stilbene compounds with varying bridging unit lengths was comprehensively characterised by Tomoda[175] and Inoue.[176] Through a combination of X-ray crystallography and computational studies, the *E* isomer of a methyl-substituted variant (*E*-**39***)* was found to be almost fully planar, whilst the ethylene bond of the *Z*-isomer (*Z*-**39**) is twisted by approximately 5.7° (Figure 2.2).[175] This isomer therefore displays helical chirality, discussed further below.

Stiff-stilbenes have already found application in the design of light-triggered supramolecular systems, particularly in the design of photoresponsive anion receptors.[172,177–180]

Wezenberg and Feringa recently reported a bis-urea functionalised stiff-stilbene derivative *E*/*Z*-**40** for which the affinity for acetate and phosphate anions could be tuned by reversible photoisomerisation between the *E* and *Z* forms. Indeed, the association constant for phosphate of the *Z* state (2.0 x 10^3 M^{-1}) was approximately 25 times greater than the molecule in the *E* form (Figure 2.3).[178] In a later study considering binding to chiral anions, the authors found that the two helical forms of the *Z*-isomer interconvert rapidly (on the NMR timescale) at room temperature, and temperatures below -55 °C were required to resolve the resonances of the diastereomeric complexes of the two helices with the single enantiomer of a BINOL-derived phosphate salt. In agreement with the experimental observations, the barrier to interconversion of the stiff-stilbene *Z*-**32** helical forms was calculated by DFT to be 16.7 kJ mol^{-1}.[181]

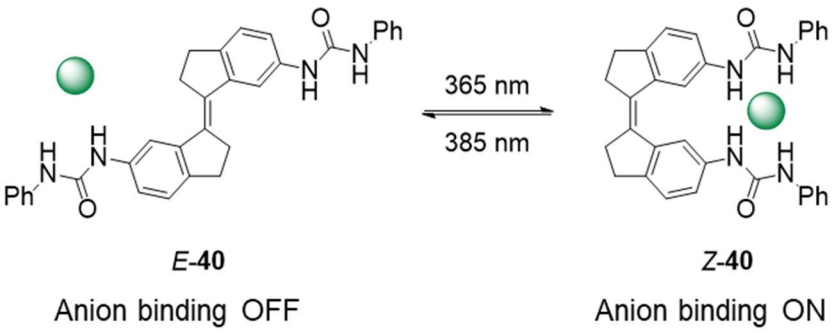


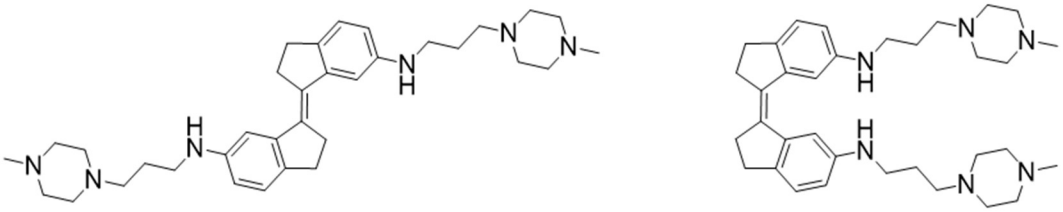
Figure 2.3: Previous applications of a stiff-stilbene derivative in the design of a photoresponsive anion receptor.

Recently, Feringa and co-workers incorporated a hindered stiff-stilbene-based molecular motor into a duplex DNA hairpin in order to allow the photocontrol of duplex stability.[182] However, the use of stiff-stilbenes as the basis of selective DNA-targeting molecules that may be eventually applied in biological circumstances, such as photopharmacology (Section 1.4.3) has not yet been explored. Therefore, as an initial proof-of-concept, the first goal of this project was to design and synthesise a set of stiff-stilbene based DNA ligands and evaluate the potential of the *E* and *Z* forms to selectively target G4 DNA in physiologically-relevant conditions.

2.2.3 Synthesis of a proof-of-concept ligand pair

As a first step to evaluate the potential of photoresponsive stiff-stilbene compounds as G4-targeting molecules, a prototype ligand pair was required that featured appropriate substituents to confer aqueous solubility and potent DNA-binding properties to the

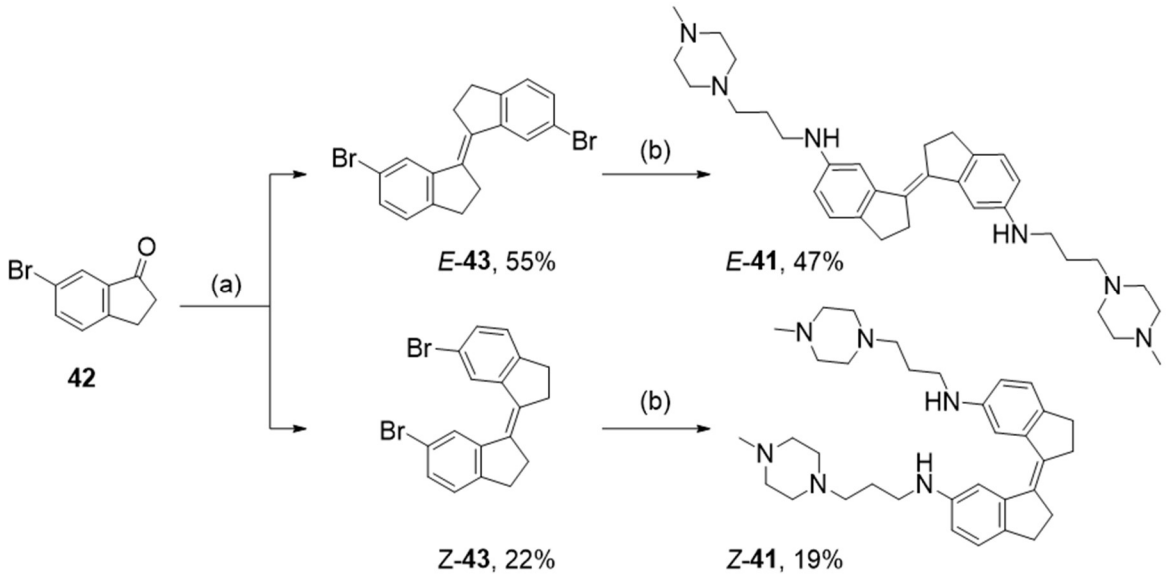
photoresponsive scaffold. A common approach in G4 ligand design that simultaneously addresses both these issues involves the tethering of basic amine moieties to the aromatic scaffold using flexible linking chains (e.g. as in the case of NDI derivative MM41 **9**, Figure 1.14). As well as rendering the ligand soluble in water, these groups are sufficiently basic (pKa ≈ 9)[183] to be protonated under physiological conditions (pH = 7.4). The resulting cationic moieties participate in electrostatic interactions with the negatively charged DNA phosphate backbone which complements the π-stacking of the ligand chromophore and G-tetrads. The flexible orientation of these pendant groups afforded by the spacer unit allows them the freedom to interact with loop and groove residues. Previously, methyl piperazine moieties have been demonstrated as effective for conferring excellent G4 recognition properties to naphthalene diimide scaffolds,[122,124] with spacer units of two or three atoms in length being sufficiently long to allow the capping amines to target groove and loop residues whilst the aromatic component participates in end-stacking interactions with the G-tetrads. This precedent pointed to the design of ligands *E*-**41** and *Z*-**41** as an initial pair to consider the general DNA-recognition ability of the stiff-stilbene scaffold and also examine the effect of the structural difference between the *E*/Z forms on DNA-binding activity (Figure 2.4).



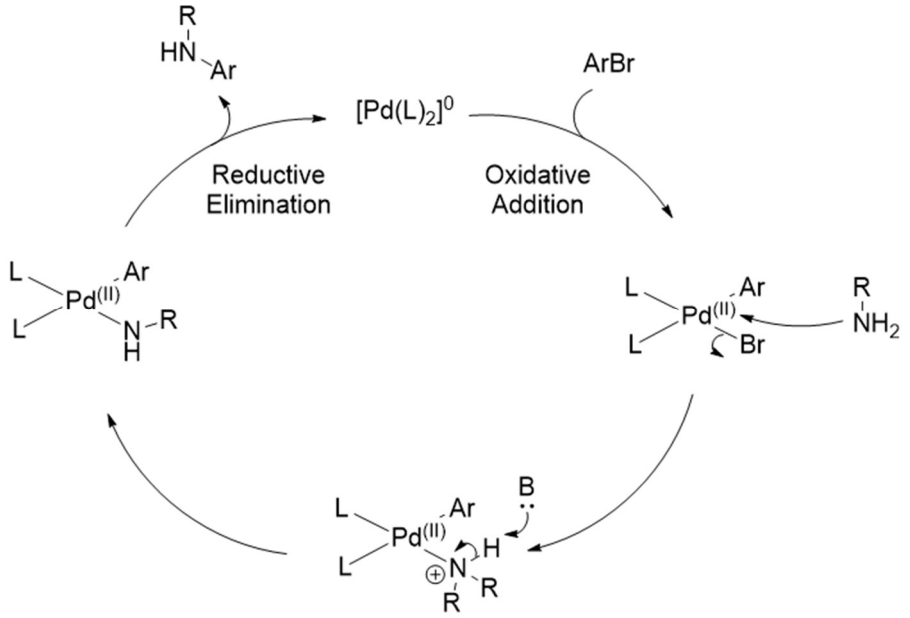
E-**41** *Z*-**41**

Figure 2.4: Prototype stiff-stilbene G4 ligands to investigate the ligand design hypothesis.

The synthesis of ligands *E*-**41** and *Z*-**41** is shown in Scheme 2.1. The preparation of the bis-bromide derivatives *E*-**43** and *Z*-**43** from 6-bromo-1-indanone **42** had previously been reported, employing the aforementioned McMurry coupling reaction for the formation of the hindered tetrasubstituted double bond.[178] This reaction proceeded in 77% overall yield to afford *E*-**43** and *Z*-**43** as a 2:1 mixture, which were separated by precipitation of the less-soluble *E*-isomer. With bis-bromides *E*-**43** and *Z*-**43** in hand, the target compounds were furnished by Buchwald-Hartwig C-N cross coupling of these intermediates with commercially available 3-aminopropyl-4-methylpiperazine. An initial attempt to employ $Pd(OAc)_2$ as the catalyst was unsuccessful, but a system comprising BrettPhos[184] Pd G3 in dioxane with sodium *tert*-butoxide as the base generated the desired products *E*-**41** and *Z*-**41**.



Scheme 2.1: Synthetic route to prototype stiff-stilbene ligands *E*-**41** and *Z*-**41**. Reagents and conditions: (a) Zn, $TiCl_4$, THF, 16 h, 65 °C; (b) 3-aminopropyl-4-methylpiperazine, BrettPhos Pd G3 (5-10 mol%), NaO^tBu, dioxane, 2 h, 80 °C.



Scheme 2.2: General mechanism of Buchwald-Hartwig C-N cross-coupling. L = ligand, B = base.

The Buchwald-Hartwig amination proceeds *via* a Pd(0) catalytic cycle (Scheme 2.2).[185] Oxidative addition of the aryl bromide coupling partner to the metal centre generates a Pd(II) species. Displacement of the bromide ligand by the amine coupling partner followed by

deprotonation by the base generates the final Pd(II) intermediate, from which reductive elimination forms the arylated amine product and regenerates the active Pd(0) species which can participate in further catalytic cycles. The products were purified by HPLC as their trifluoroacetate salts and characterised by NMR spectroscopy.

2.2.4 Synthesis of a compound library

Whilst the prototype ligand couple *E/Z*-**41** would allow an initial assessment of the potential differences in G4 recognition between the isomeric forms of the stiff-stilbene scaffold, further analogues were desired to elicit possible structure-activity relationships that may be present in the ligand series. Several points of difference were considered:

(a) variation in the length of the spacer unit, in order to determine the optimal length of this component: spacers of three to seven atoms in length were considered.

(b) variation in the chemical functionality of the aryl substituent, in order to examine the electronic influence of the aryl substituent and the effect of increased rigidity in the spacer unit: the effect of substituting the amine linkage for a urea linkage was explored.

(c) variation in aryl substitution pattern, noting that the electronics of the compound are different depending on the position of the electron-donating nitrogen substituent relative to the central stilbene moiety: 5-substituted (*para* to stilbene olefin) and 6-substituted (*meta* to stilbene olefin) derivatives were considered.

(d) variation in the nature of the terminal basic moiety, since previous studies have shown that this can have a marked effect on G4 binding properties: the difference between methyl piperazine and morpholine termini was investigated.

Figure 2.5 summarises the compound library synthesised to investigate the above effects.

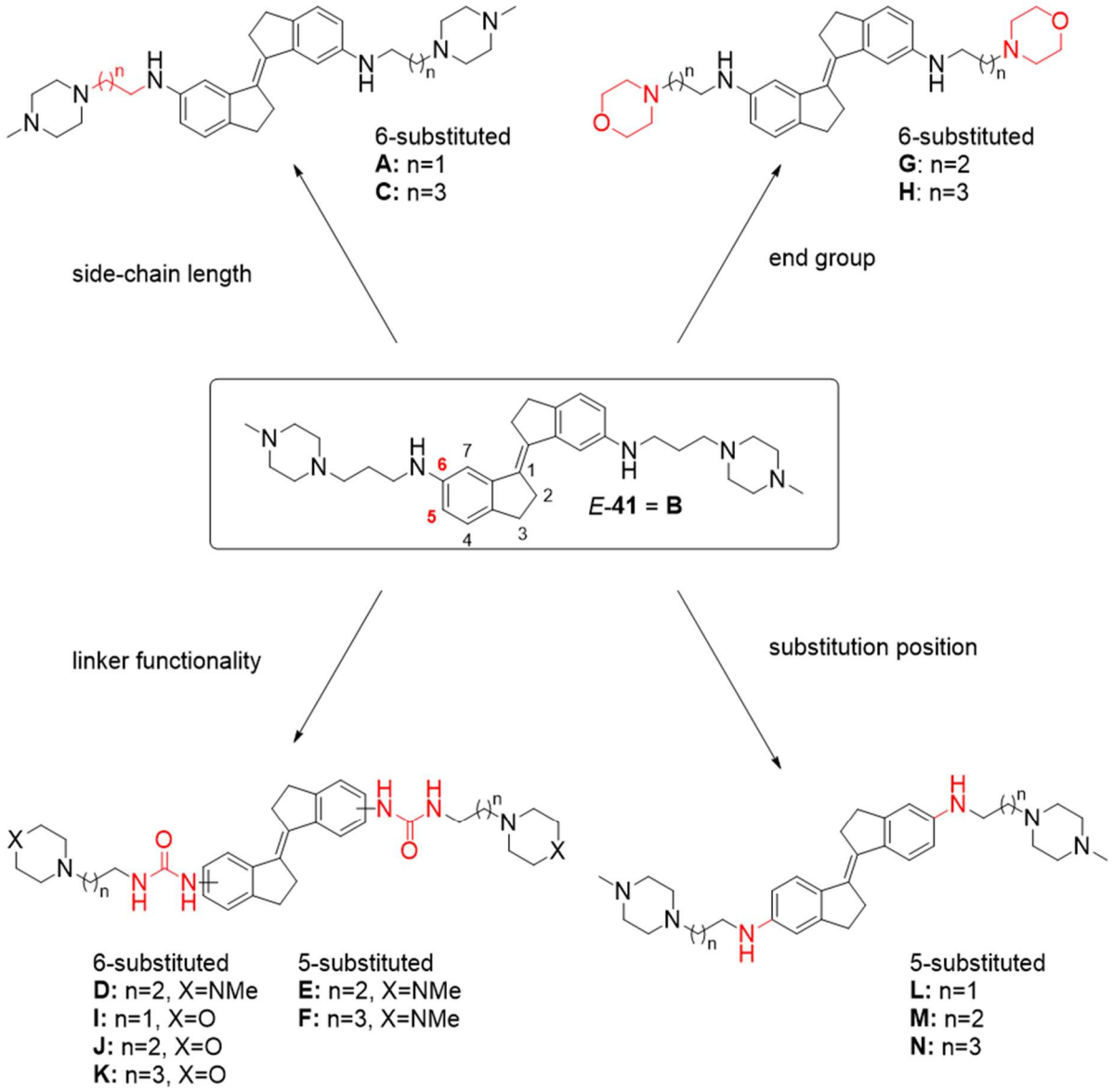


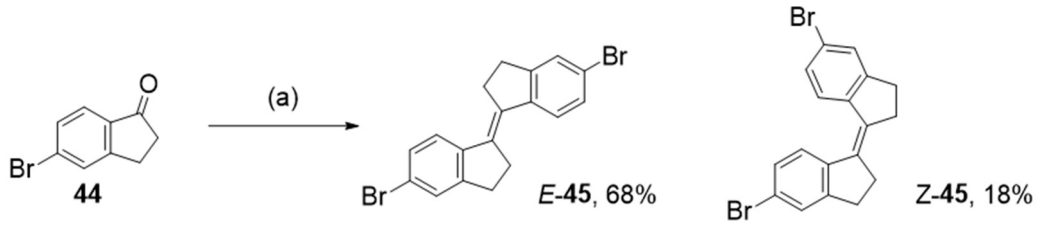
Figure 2.5: Overview of stiff-stilbene compounds synthesised, showing the various points of structural difference from *E*-**41**. The compounds in the library are identified **A-N,** with compound *E*-**41** identified as **B** in this series.

Preparation of the 5-substituted isomers required bis-bromide *E*-**45**, which was prepared using the same procedure used for the synthesis of the 6-substituted bromide (Scheme 2.2) from 5-bromo-1-indanone **44**. As before, a mixture of *E* and *Z* isomers were formed, and the *E* isomer separated by precipitation. Though this isomer could reasonably be expected to be the less soluble isomer due to its planarity, the stereochemistry of the products was verified by comparing the NOESY spectra of the two isomers, which confirmed the spatial proximity of

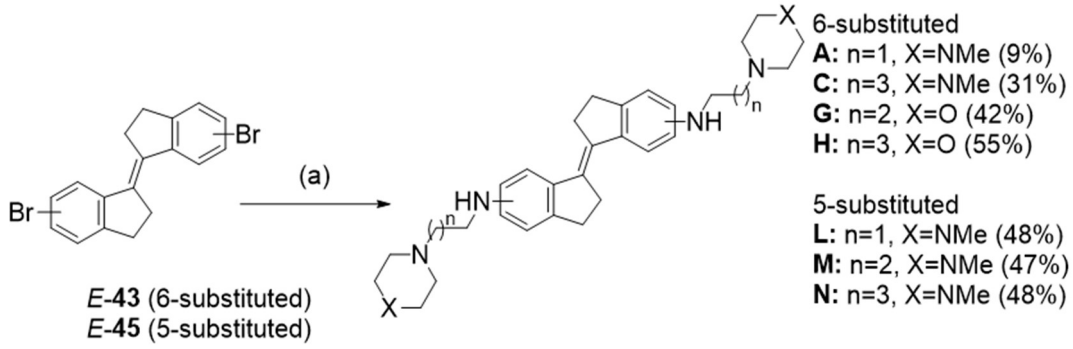
the aromatic H4 and H7 hydrogen atoms and bridging methylene groups in the *E* isomer (Figure 2.6).

With this key intermediate in hand, the amine-linked derivatives were readily accessed using the same Buchwald-Hartwig approach used for the synthesis of *E*-**41** using the appropriate bromide and commercially-available primary amine (Scheme 2.3).

Unfortunately, though 5-substituted derivatives **L**, **M** and **N** were successfully synthesised, isolated and characterised, they began to decompose in aqueous solution overnight (Appendix, Figure A2.1) and were therefore not considered suitable candidates for biophysical screening. This unwanted reactivity was rationalised to arise from activation of the stilbene double bond by the mesomeric electron donating effect of the *para* amine substituent. Fortunately, decomposition was not observed for the other analogues, even after several days in aqueous solution at ambient temperature.



Scheme 2.2 Synthesis of novel bis-bromides *E*/*Z*-**45**. Reagents and conditions: (a) Zn, $TiCl_4$, THF, 30 h, 65 °C.



Scheme 2.3: Synthesis of amine-linked stiff-stilbene G4 ligands **A**,**C**,**G**,**H**,**L**,**M**,**N**. Reagents and conditions: R-NH_2, BrettPhos Pd G3 (5-10 mol%), NaO^{t}Bu, dioxane, 2 h, 80 °C.

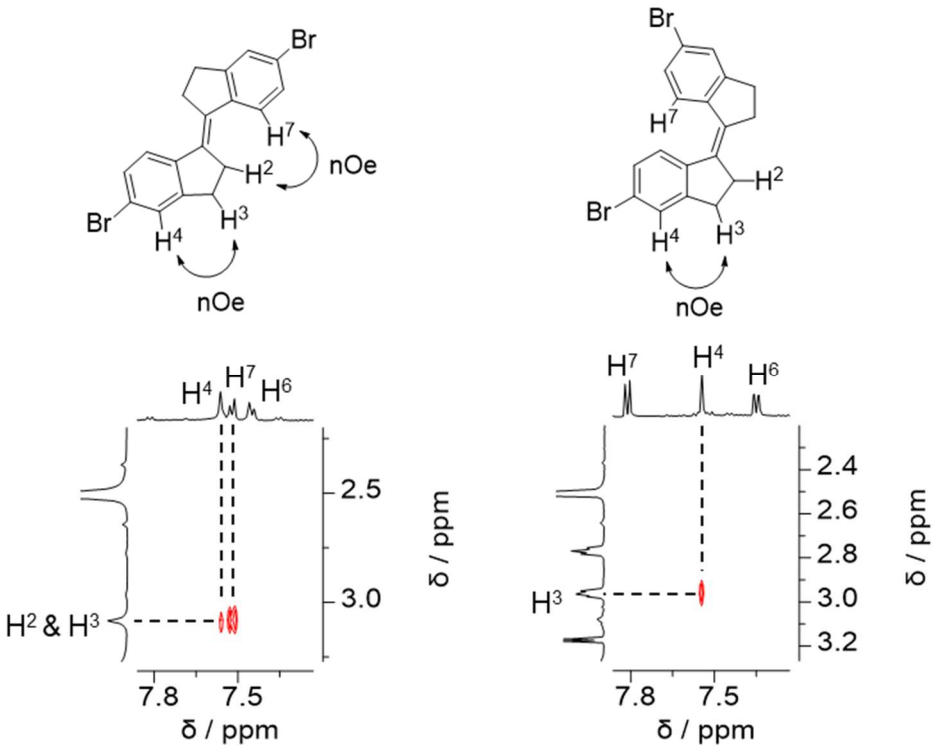
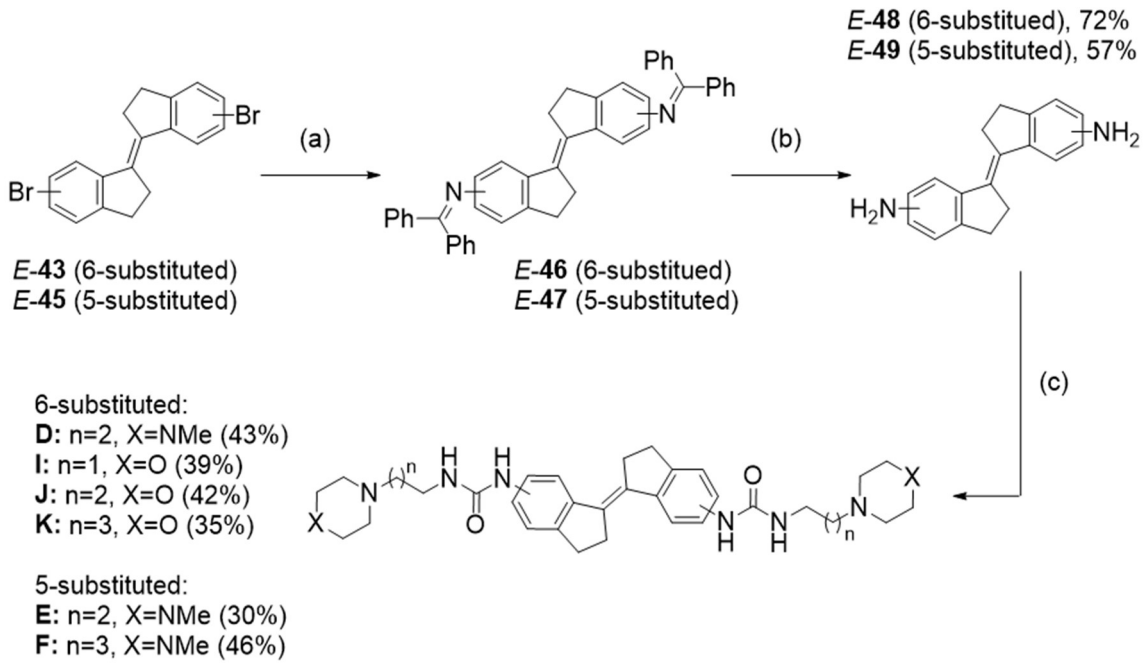


Figure 2.6: Assignment of E/Z stereochemistry in novel bromides by nOe analysis. The isomers are differentiated by the absence of nOe between H7/H2 in the *Z*-isomer.

Meanwhile, the urea derivatives were accessed using the procedure shown in Scheme 2.4. Buchwald-Hartwig amination of the appropriate bromide (*E*-**43** or *E*-**45**) with benzophenone imine following a previously-reported procedure[178] afforded intermediates *E*-**46** and *E*-**47,** which are subsequently hydrolysed under acidic conditions to furnish amines *E*-**48** and *E*-**49** in good yield (57-72%) over two steps. Activation with triphosgene followed by reaction with the appropriate amine afforded the required urea derivatives in yields of 30-46% following purification by reverse-phased HPLC.



Scheme 2.4: Synthesis of urea-linked stiff-stilbene G4 ligands **D**,**E**,**F**,**I**,**J**,**K**. Reagents and conditions: (a) benzophenone imine, $Pd(OAc)_2$ (10 mol%), DPPF (15 mol%), NaO*t*Bu, PhMe, 90 °C; (b) 2 M HCl (aq), THF, r.t.; (c) triphosgene, Et_3N, DCM, r.t., then R-NH_2.

2.3 Biophysical studies

With the prototype ligand pair (*E*-**41** and *Z*-**41**) and the rest of the compound library (**A**-**H**) in hand, efforts turned to evaluating the efficacy of these compounds as G4 binding ligands through a variety of biophysical screening assays.

2.3.1 FRET thermal melting assays

2.3.1.1 Introduction to FRET

The fluorescence resonance energy transfer melting assay is an invaluable means to rapidly screen small molecules for interaction with biological targets.[108] This assay measures the effect of a small molecule on the thermal stability of the biomolecular secondary structure. Ligand binding is an energetically favourable process, meaning there is a greater energetic penalty for the denaturation process and higher temperatures are needed to induce unfolding. In this case, the biomolecule of interest is obtained conjugated to a fluorescent donor at one end and a fluorescence acceptor at the other. The FRET pair is chosen such that the

absorption of the donor does not overlap with that of the acceptor. The donor can then be excited specifically, and since the emission spectra of the two species are also distinct, the fluorescence of only one species can be monitored. The well-established FAM (**50**)/TAMRA (**51**) pair is most commonly employed in the design of fluorescent nucleotides for G4 studies (Figure 2.7). In the absence of the TAMRA acceptor, excitation of the FAM donor at its absorbance wavelength (495 nm) results in fluorescence emission at longer wavelength (520 nm). However, if the acceptor is present in close proximity, the resonance between the donor emission and acceptor excitation allow non-radiative energy transfer to the acceptor, effectively quenching the fluorescence of the donor (520 nm) and resulting instead of the fluorescence of the acceptor (585 nm).

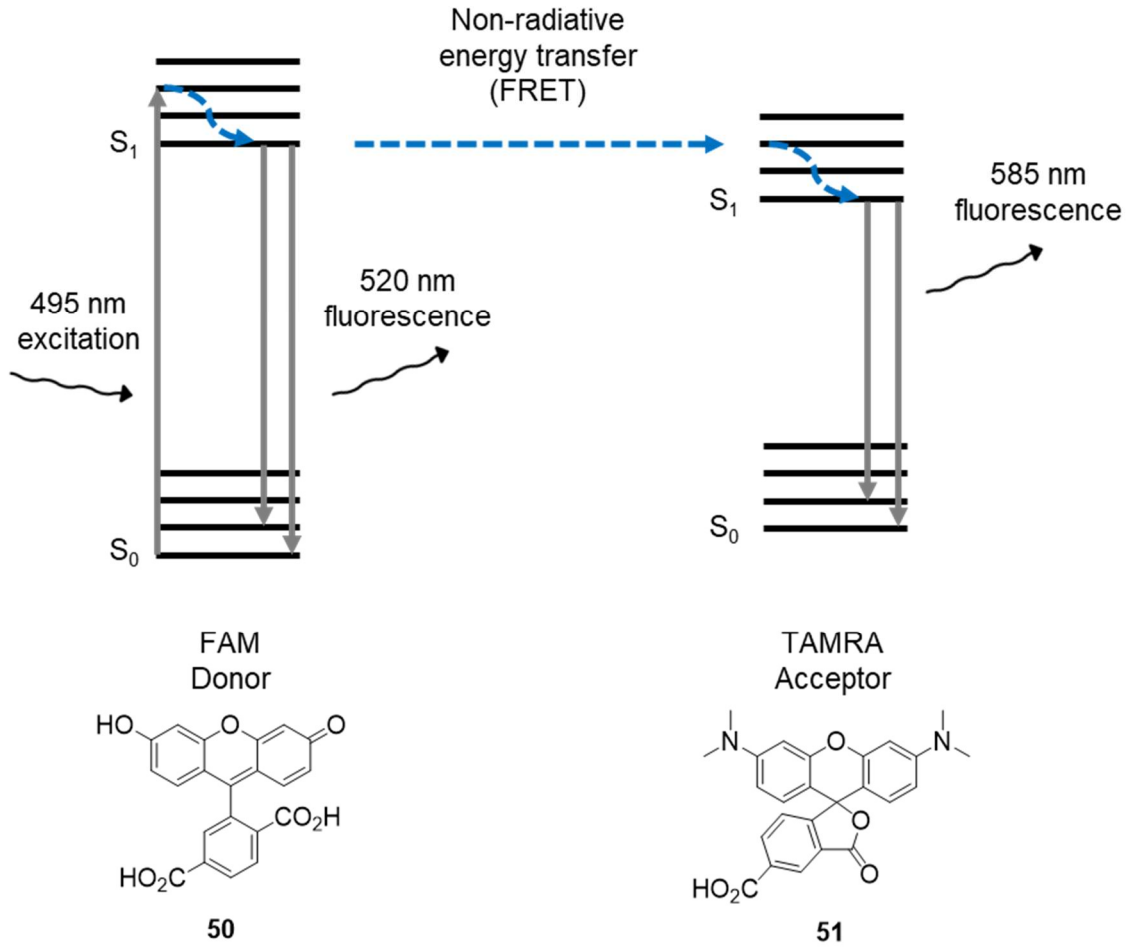


Figure 2.7: Jablonski diagram demonstrating FRET in the FAM/TAMRA fluorescence pair.

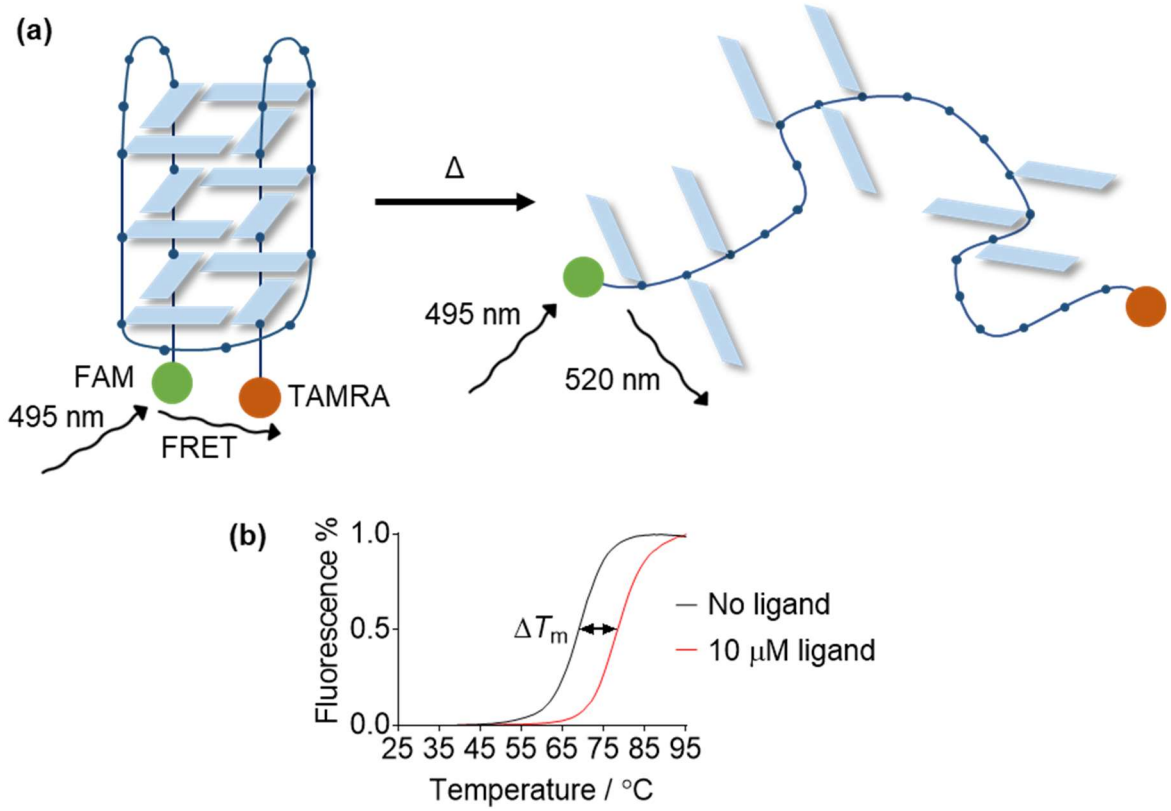


Figure 2.8: Schematic overview of the FRET melting assay. (a) thermal denaturation of the secondary structure causing the fluorophore pair to move further apart in space, leading to increased donor fluorescence as the temperature is increased; (b) example melting curves and representation of ΔT_m.

Since the efficiency of FRET depends strongly on the distance between the acceptor and donor, monitoring the donor fluorescence therefore allows the denaturation of the secondary structure to be followed (Figure 2.8a). In the folded form of the biomolecule the fluorophores are in close proximity leading to efficient quenching of donor (FAM) fluorescence. As the secondary structure denatures upon raising the temperature the fluorophores move apart, resulting in an increase in observed fluorescence. Normalisation of this signal and determination of the temperature at which the donor fluorescence is half of the maximum indicates the characteristic melting temperature of the secondary structure (T_m), which can be compared in the presence and absence of ligand (ΔT_m) to indicate the ability of the ligand to bind to and stabilise the folded structure (Figure 2.8b).

Compared to the other biophysical techniques available during this project for the study of G4/ligand complexation, the FRET assay has several distinct advantages. Though the melting

of the secondary structure can be followed by a range of other techniques (e.g. UV/visible absorbance spectroscopy), fluorescence is particularly useful since the sensitivity of this technique only requires very small quantities of the biomolecule. Furthermore, the assay is readily amenable to the 96-well plate format, allowing a large number of compounds to be screened rapidly against a panel of DNA sequences. However, a disadvantage of the method is that the introduction of the artificial fluorophores into the oligonucleotide models may lead to the generation of artefacts, for example by interaction of the ligand with the components of the FRET pair rather than the oligonucleotide. Furthermore, since the assay is temperature dependent, the association of the ligand for the biomolecule under physiological conditions cannot be quantified, since this is in itself a temperature-dependent parameter.[186] Therefore, whilst the assay provides a rapid means to screen molecules for binding to DNA structures, promising compounds must later be validated using further techniques to gain confidence in the results.

2.3.1.2 Library screen

The ability of all *E*-configuration compounds (**A**-**H**) to induce thermal stability in G4 and duplex DNA was first assessed, in order to provide an initial assessment of the potential of the stiff-stilbene scaffold to serve as the basis of DNA-binding agents. For this initial screen the human telomeric sequence (F21T) in potassium-containing buffer was selected as the G4 model, owing to the physiological relevance of this G4 sequence and that this model is almost ubiquitously employed in the G4 ligand binding studies previously reported.[108] For the duplex model, a self-complementary hairpin (F10T) comprising ten base pairs separated by a hexaethyleneglycol (HEG) linker was employed, being of comparable size to the G4 sequence.[122] Compounds were screened at 5 μM and 10 μM ligand concentration in order to identify meaningful hits; compounds that do not display induce significant thermal stabilisation at these concentrations are not considered effective ligands. The results are shown in Table 2.1 and representative raw data in the Appendix A2 (Figures A2.2-A2.12).

Table 2.1: FRET thermal stabilisation values (ΔT_m) induced by ligands **A**-**K** in G4 (F21T-K^+) and duplex (F10T) sequences. DNA concentration = 200 nM. Further experimental details are provided in Section 7.3.1.

	ΔT_m / °C			
	F21T-K^+ G4		F10T duplex	
Ligand	10 μM	5 μM	10 μM	5 μM
A	12 ± 0.9	5 ± 0.5	0 ± 0.1	0 ± 0.1
B (*E*-**41**)	12 ± 0.5	6 ± 1.1	0 ± 0.1	0 ± 0.2
C	9 ± 1.2	5 ± 1.8	0 ± 0.3	0 ± 0.4
D	11 ± 0.8	4 ± 0.3	0 ± 0.2	-1 ± 0.1
E	8 ± 0.7	1 ± 0.3	0 ± 0.1	-1 ± 0.1
F	7 ± 0.5	1 ± 0.5	0 ± 0.1	-1 ± 0.2
G	-4 ± 0.7	-8 ± 0.3	-1 ± 0.1	-1 ± 0.2
H	-2 ± 0.6	-6 ± 0.4	-1 ± 0.1	-1 ± 0.1
I	-4 ± 0.7	-5 ± 0.1	-1 ± 0.1	-1 ± 0.1
J	-4 ± 0.5	-5 ± 0.2	-1 ± 0.1	-1 ± 0.2
K	-4 ± 0.5	-5 ± 0.8	-1 ± 0.1	-1 ± 0.1

Pleasingly it was observed that all compounds bearing pendant methylpiperazine moieties (**A**-**F**) induced significant thermal stabilisation (ΔT_m = 7-12 °C) in the G4 model at 10 μM ligand concentration (Table 2.1). Whilst stabilisation of the G4 structure remains evident at 5 μM ligand concentration, the stabilisation temperatures at this concentration are rather weak (ΔT_m = 1-6 °C), which suggested that further optimisation of this ligand series was likely required to improve G4 affinity for the purposes of biological applications. However, it should be noted that the stabilisations achieved at 10 μM are higher than those reported for previously reported azobenzene derivatives (ΔT_m 1-3 °C) studied at higher (40 μM) concentration.[167] This comparison suggests that the high conformational rigidity and large hydrophobic surface area of the stiff-stilbene scaffold is an important feature that furnishes this scaffold with good G4-targeting properties.

Within the methylpiperazine compound class, the effect of ligand structural variation of G4 stabilisation appears to be rather modest, though some tentative relationships can be inferred. Compounds with shorter spacer units appeared to stabilise G4 slightly more strongly, for

example compound **B**, featuring a 4-atom spacer confers higher thermal stability (ΔT_m = 12 °C) than compound **C** bearing a longer spacer unit (ΔT_m = 9 °C). Similarly, comparing the urea derivatives, **E** (containing 6-atom spacer) induces a slightly higher thermal stabilisation than **F** which features an additional methylene group in each linking chain. The substituent position of the aromatic core also appears to affect the G4 stabilisation, with compound **D** (6-substituted) displaying a higher (ΔT_m = 11 °C) stabilisation than 5-substituted analogue **E** (ΔT_m = 8 °C). This effect could result from slightly better presentation of the side-chains towards the quadruplex grooves (assuming an end-stacking binding mode for the central aromatic core) in the 6-substituted variant, or could result from the influence of the substituent on the electronics of the stilbene core, since the nitrogen lone pair is in conjugation with the central stilbene double-bond in the 5-substituted isomers. It should be noted that whilst these minor structure-activity effects are discernible they are not sufficiently striking, nor the compound library sufficiently large, to make wider inferences about the optimal features of methylpiperazine-based stiff-stilbene ligands. Whilst the synthesis and evaluation of additional analogues would provide more convincing evidence for some of the structure-activity relationships emerging in this compound class, efforts towards this goal were not pursued because, at least based on the results obtained in this initial screen, these relatively minor modifications were unlikely to lead to the dramatic improvement in G4 binding activity that was ultimately sought.

Whilst only small changes in activity appear to result from the structural modifications in the methylpiperazine compound series, the nature of the terminal basic residue appears to exert a much more dramatic effect on the interaction of these stiff-stilbene derivatives with G4 DNA. All compounds in the morpholine series (**G**-**K**) appear to induce a reduction in the thermal stability of the G4, of ~ 5 °C, relative to the melting temperature of the G4 in the absence of ligand. This surprising effect has also been observed by others but never fully investigated.[33] Whilst further exploration of this apparent destabilisation was not undertaken during the present project, further studies are warranted to understand the true origins of this effect.

2.3.1.3 G4/duplex selectivity

Critically, no stabilisation of the duplex DNA hairpin is observed for any of the compounds, indicating that the stiff-stilbene scaffold is effective at discriminating against duplex DNA binding in favour of targeting G4. To verify the selectivity of the scaffold against duplex, the lead compound **B** (*E*-**41**) was investigated in a competition assay, in which increasing concentrations of a non-fluorescent duplex DNA sequence are added to the solution of the G4/ligand complex. In this case, competitive off-target binding to the duplex sequence diminishes the concentration of the G4/ligand complex, observed as an apparent depression

in the induced thermal stabilisation of the fluorescent G4 species (Figure 2.9). Though a gradual depression of the ΔT_m caused by off-target binding is evident, indicating the G4/duplex discrimination is not perfect, the thermal stabilisation remains significant (ΔT_m = 9 °C) even in the presence of 10 μM duplex DNA competitor, corresponding to 50 equivalents of duplex hairpins versus quadruplex (0.2 μM) strands. That 75% of the G4 thermal stabilisation potential of the ligand is retained in the presence of a high concentration of duplex competitor demonstrates the preference of the ligand for the four-stranded structure that corroborates the initial result of the initial screen, where no stabilisation of the duplex strand was observed.

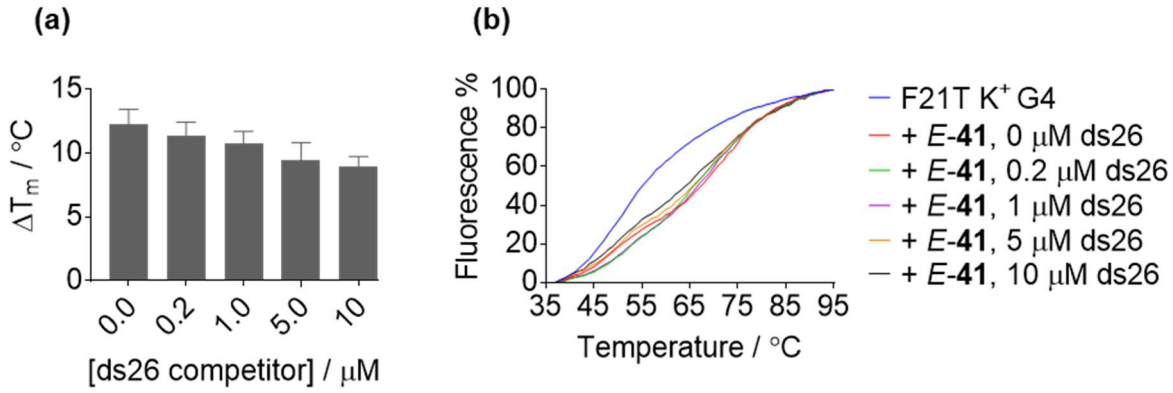


Figure 2.9. FRET competition assay examining the effect of increasing concentrations of non-fluorescent duplex competitor (ds26) on ***E*-41** induced thermal stabilisation of F21T-K^+ G4.

*2.3.1.4 Difference in activity of E-**41** and Z-**41** stiff-stilbenes*

Having demonstrated methyl-piperazine decorated stiff-stilbene ligands can selectively target G4 DNA in the initial screen of the compound library, efforts moved towards exploring possible differences in the *E*/*Z*-**41** ligand pair on G4 recognition in pursuit of the eventual goal of photoswitchable control of the ligand activity. Since ***E*-41** already demonstrated affinity for G4 in the initial screen, a wider panel of G4 structures were included in the comparative study of these two compounds. Compounds were also screened at concentrations down to 1 μM to provide a clearer indication of their potency at low micromolar concentrations. In addition to the F21T telomeric G4 model in potassium-rich buffer previously discussed, ***E*-41** and ***Z*-41** were screened against the same sequence in sodium-containing buffer, since telomeric G4 is observed to form an alternative antiparallel G4 topology under these conditions.[27] Two further G4s were included relevant to the ultimate potential therapeutic applications of the ligands. A sequence corresponding to the G4 found in the promoter region of the *c-myc* oncogene was

included, as a model of a further potential anticancer target (see Section 1.3.1.2).[187] Finally, a G4 sequence found in *T. brucei* parasite genome was studied, owing to the potential of G4 ligands to serve as antiparasitic agents (see Section 1.3.1.3). The detailed structure of this sequence is not fully characterised, but preliminary CD experiments by Belmonte-Reche *et al.* suggest it exists as a mixture of G4 topologies under assay-relevant conditions.[52]

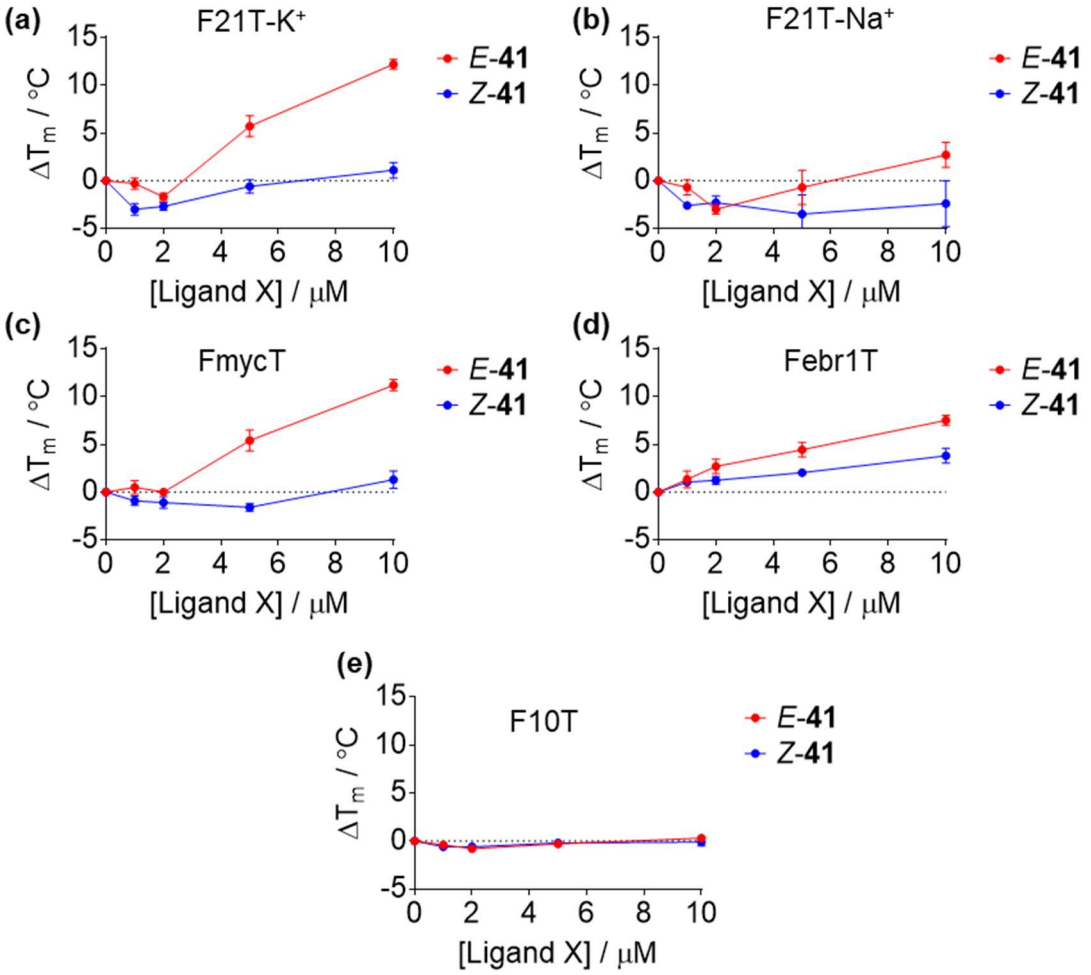


Figure 2.10. Concentration-dependent FRET thermal stabilisation values (ΔT_m) induced by ligands *E*-**41** and *Z*-**41** in G4 and duplex sequences: (a) F21T-K^+; (b) F21T-Na^+; (c) FmycT; (d) FebrT; (e) F10T.

The results of the study are shown in Figures 2.10 and representative raw data in Appendix A2 (Figures A2.12-2.13). The most striking feature of the results is that the isomeric configuration of the stiff-stilbene core exerts a large effect on the activity of the ligand. Indeed,

for the F21T-K^+ (Figure 2.10a) and FmycT G4 (Figure 2.10c), whilst *E*-**41** induces thermal stability of up to 12 °C for both sequences at 10 μM concentration, *Z*-**41** is ineffective and does not induce any effect on G4 stabilisation over the concentration range examined. The thermal melting curves for F21T-K^+ clearly show this difference in activity (Figures 2.11a and 2.11b). The curves obtained for the experiments with F10T duplex are also shown, indicating no stabilisation of the hairpin structure by either ligand (Figures 2.11c and 2.11d). The positive shift in the thermal melting curves induced by ligand *E*-**41** is clearly visible whilst the G4 melting curve remains comparatively unperturbed by ligand *Z*-**41**. *E*-**41** also appears to stabilise the Ferb1T quadruplex more effectively than *Z*-**41**, although the difference in activity is less pronounced in this case (Figure 2.10d). The difference in activity observed between the *E*-**41** and *Z*-**41** appears to validate the design hypothesis which proposed the *E* derivative should be the more active ligand, since the central aromatic core adopts a planar geometry and can therefore be expected to interact with the G4 tetrads better than the twisted *Z* isomer.

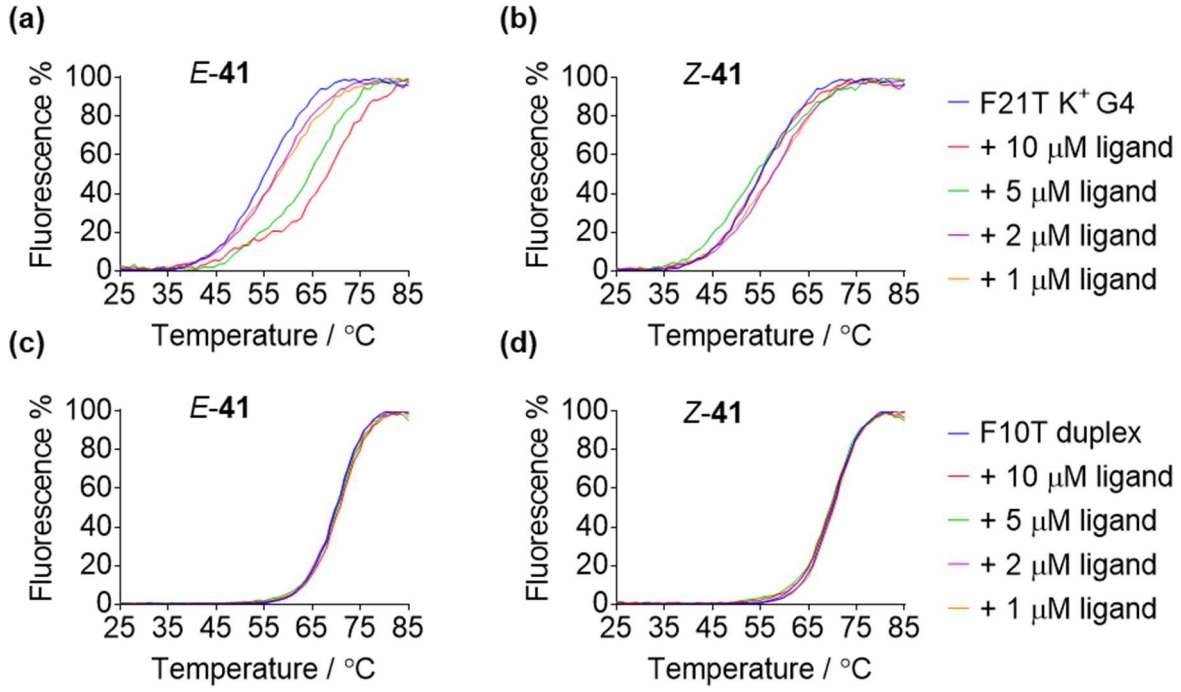


Figure 2.11. FRET thermal melting curves for (a)-(b) F21T-K^+ G4 and (c-(d) F10T duplex in the absence (blue) and presence of increasing concentrations of (a),(c) *E*-**41** and (b),(d) *Z*-**41**.

A further noteworthy feature of these results is that though *E*-**41** appears to stabilise the F21T-K^+ (hybrid), FmycT (parallel) and FebrT (mixed) G4 topologies, this ligand induces negligible stabilisation of the F21T-Na^+ (antiparallel) G4 (Figure 2.10b). This indicates that, in addition to

selectivity against duplex DNA, the stiff-stilbene scaffold may also display some discrimination between different types of G4 folded structure. This interesting observation was further investigated by UV/visible and NMR titration studies (*vide infra*).

Taken together, the results of the FRET assays demonstrated the promise of the new stiff-stilbene ligand series. The ligand design appears to result in affinity for G4 and promising binding selectivity between different DNA topologies. Critically, regarding the eventual aim to control the activity of the ligands by photoisomerisation, the two isomers display differing G4 binding activity. However, it should be noted that the FRET assay only serves to provide a preliminary insight into the association of the compounds with G4 DNA. As discussed earlier, this assay is susceptible to artefacts such as the interaction of the ligands with the fluorescent dyes, which could generate false positive or false negative results. Additionally, ligand binding may itself perturb the DNA folding topology and therefore results obtained cannot be unambiguously attributed to stabilisation of the native structure formed in the absence of ligand. Moreover, the FRET assay provides no structural information on G4/ligand complexes that would validate the binding mechanisms proposed to result in the observed activities. Therefore, further investigation of the G4 binding of the ligands was necessary to confirm the key results and provide additional information to guide further ligand design. The following sections report the more detailed investigation of the *E/Z*-**41** couple using additional techniques, including circular dichroism, UV/visible absorbance spectroscopy and NMR spectroscopy.

2.3.2 UV/visible absorbance titrations

2.3.2.1 Introduction to the technique

The electronic structure of molecules with delocalised π-systems results in characteristic UV/visible absorbance spectra. In simple terms, a photon at resonance with the energy difference between an occupied and unoccupied molecular orbital is absorbed, causing the excitation of the molecule to a higher energy state if the transition is permitted on quantum mechanical grounds.[188] The Beer-Lambert law states that the observed absorbance (at a path length of 1 cm) of a solution of species Y, A_Y, is proportional to the concentration, $[Y]$.[189] The constant of proportionality is the molar extinction coefficient at the wavelength in question, ε_Y:

$$A_Y = \varepsilon_Y \cdot [Y] \tag{2.1}$$

In the case of a titration of ligand (L) with DNA (D), complex formation ($D.L$) occurs according to the equilibrium:

$$D + L \rightleftharpoons D.L \tag{2.2}$$

The affinity of the two species, i.e. the degree of complex formation at equilibrium at a particular concentration of D and L, is quantified by the association constant (K_a),[186] defined as:

$$K_a = \frac{[D.L]}{[D] \cdot [L]} \tag{2.3}$$

In a typical UV/visible absorbance titration, the total concentration of ligand is held constant, whilst the concentration of DNA is varied. Assuming the Beer-Lambert law is valid, the absorption spectrum of a mixture is a linear combination of the spectra of the constituent species weighted by their respective concentrations. The observed spectrum ($A_{[D]}$) at a particular titration point (i.e. concentration of DNA) is therefore given by:

$$A_{[D]} = \varepsilon_D[D] + \varepsilon_L[L] + \varepsilon_{D.L}[D.L] \tag{2.4}$$

Outside of the range of DNA absorption (λ > 320 nm) ε_D is zero and the observed spectrum is simply a combination of that of the free ligand and bound ligand:

$$A_{[D]} = \varepsilon_L[L] + \varepsilon_{D.L}[D.L] \tag{2.5}$$

Complexation of the ligand and DNA is therefore detected by a change in the observed ligand absorbance spectrum upon addition of DNA, as the concentration of free ligand decreases and the bound complex forms. The spectra of the free and bound ligand are almost invariably different, since the intermolecular forces between the macromolecule and ligand perturb the ligand's electronic structure. For example, stacking interactions between the π systems of the DNA bases and aromatic ligand scaffolds decrease the energy gap between the π and π* ligand molecular orbitals, resulting in a bathochromic (red) shift in the absorbance spectrum.

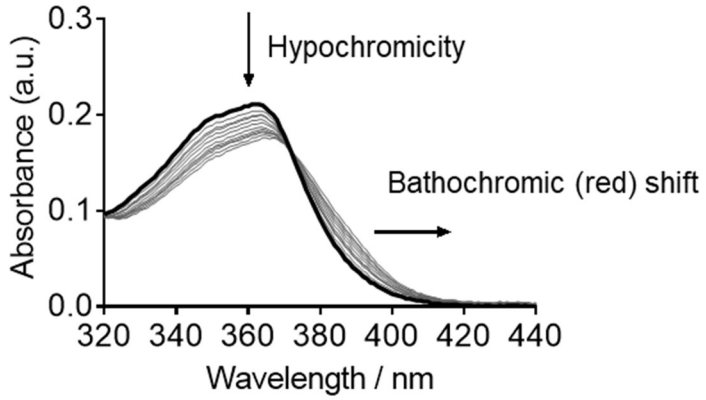


Figure 2.12. Example ligand UV-titration spectrum showing spectral changes on complexation with DNA.

Binding to DNA may also induce hyperchromic or hypochromic shifts owing to changes in the ligand structure on binding (Figure 2.12).[190] UV/visible absorbance titrations therefore allow the formation of DNA/ligand complexes to be investigated without the requirement to modify the biomolecule with artificial fluorophores. Furthermore, unlike the FRET melting assay, the temperature is kept constant and therefore the association constant under physiologically-relevant conditions may be determined. Specifically, the change in absorbance may be expressed in terms of the association constant through Eqn. (2.6) (derived in the Appendix A1):[186]

$$\Delta A = \varepsilon_{\Delta D.L} \cdot \frac{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right) - \sqrt{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right)^2 - 4 \cdot [D]_{tot}[L]_{tot}}}{2} \qquad (2.6)$$

Values for $\varepsilon_{\Delta D.L}$ and K_a may be therefore obtained by monitoring the change in absorbance of a solution of fixed total ligand concentration on increasing the concentration of DNA and fitting the observed absorbance changes to Eqn. (2.6) using non-linear regression.

The binding of *E*-**41** and *Z*-**41** to the hybrid (telo23-K^+) and antiparallel (telo22-Na^+) forms of telomeric DNA were investigated by monitoring the absorbance changes of the ligands upon titration with the relevant G4 model. The changes to the absorbance spectra observed during the titrations are displayed in Figure 2.13.

2.3.2.1 Titrations with hybrid (telo23-K^+) G4

Comparing first the binding of ligands *E*-**41** and *Z*-**41** to hybrid telo23-K^+ G4, clear differences can be observed between the two ligand systems (Figures 2.13a and 2.13c). Upon increasing the concentration of telo23 during the titration, binding of the ligand *E*-**41** is clearly evidenced by a 15% hyperchromicity in the 365 nm absorbance maximum and a bathochromic shift of approximately 7 nm (Figure 2.13a). The emergence of such red-shifted features is suggestive of π stacking between the ligand and G4 bases. Though it is not possible to draw definitive conclusions about the binding mode, the interaction between *E*-**41** and hybrid telo23 DNA observed in the FRET assay is clearly confirmed.

In contrast, the addition of telo23 to a solution of *Z*-**41** barely perturbs the ligand spectrum (Figure 2.13c). Indeed, a weaker hypochromicity (< 10%) and bathochromic shift (3 nm) are observed. Plotting the isotherm at 390 nm indicates the difference in affinities of *E*-**41** and *Z*-**41** for the G4 structure (Figure 2.13e). The curvature observed in the *E*-**41** isotherm indicates that the complete binding of this ligand to the G4 structure is approached over the concentration range studied. Meanwhile, the small perturbations to the *Z*-**41** spectrum vary linearly over the course of the titration, indicating that that saturation of ligand binding is not approached and that ligand has weaker affinity, as previously indicated the FRET assays.

The hyperbolic isotherm was well-described by the 1:1 binding model and the blue curve in Figure 2.13e is the result of fitting the observed data points to Eqn. (2.6) by non-linear regression. This allowed the association constant to be estimated as (9.1 ± 1.7) x 10^4 M^{-1}. In comparison with the association constants of previously optimised G4 ligands this affinity is not particularly high, and association constants of 10^6 or greater are reported to be necessary for therapeutically-relevant G4 ligands.[78] However, the value obtained is in the range expected from the FRET results, where ligand concentrations comparable to those employing the UV/visible titration studies (10 μM) were required to observe stabilisation of the G4 structure. Because of the very small perturbations of the UV/visible spectrum of *Z*-**41** by telo23-K^+, the ligand dissociation constant could not be measured at this concentration range. In order to measure the dissociation constant for this ligand, the concentration must be increased in order that binding saturation is observed over the course of the titration. However, given the large quantities of DNA required for such experiments this was not pursued.

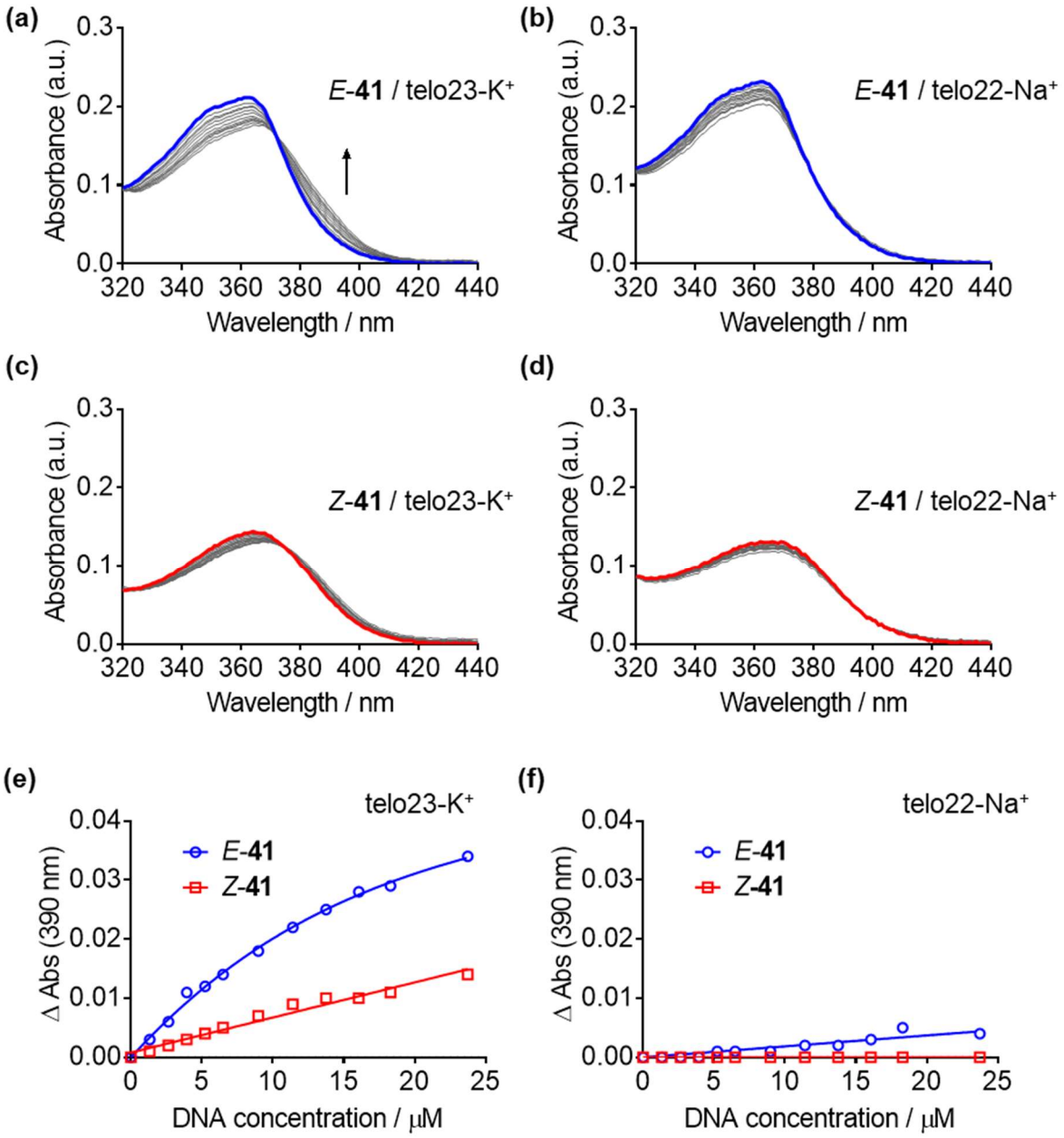


Figure 2.13: UV/visible titration studies of ligands *E*-**41** and *Z*-**41** with G4 DNA. (a)-(d): titrations of 10 μM (a),(b) *E*-**41** and (c),(d) *Z*-**41** with (a),(c) hybrid (telo23-K^+) G4 and (b),(d) antiparallel (telo22-Na^+) G4; (e)-(f): extracted 390 nm binding isotherms for *E*-**41** and *E*-**41** with (e) hybrid (telo23-K^+) and (f) antiparallel (telo22-Na^+) G4.

Table 2.2: Summary of spectral changes observed for ligands *E*-**41** and *E*-**41** (10 μM) after addition of 2.5 equiv. G4 DNA.

	Hypochromic shift / %		Bathochromic shift / nm	
	E-**41**	*Z*-**41**	*E*-**41**	*Z*-**41**
Hybrid (telo23-K^+)	15	9	7	3
Antiparallel (telo22-Na^+)	11	8	1	0

2.3.2.2 Titrations with antiparallel (telo22-Na^+) G4

Titrations with the antiparallel G4 model (telo22-Na^+) were conducted over the same concentration range. Under these conditions, though the hypochromic shifts observed are similar (though slightly lower) than those seen upon titrations with the hybrid G4 model, a significant difference is the absence of the bathochromic shift upon addition of telo22 (Figures 2.13b and 2.13d). This can clearly be observed for the active *E*-**41** isomer by inspection of the titration curves for the two quadruplexes (Figures 2.13a and 2.13b) and also by comparison of the 390 nm isotherms (Figure 2.13f) with those obtained from the telo23 titrations (Figure 2.13e). Therefore, it appears that the hybrid G4 offers a specific binding mode for ligand *E*-**41** that is comparatively inaccessible in the antiparallel form. This result appears to confirm the selectivity observed in the thermal melting assay, which indicate *E*-**41** preferentially stabilises hybrid G4 (F21T-K^+) over the antiparallel (F21T-Na^+) model. The comparable hypochromic shifts may arise from non-specific electrostatic interactions with the external regions of the G4 that are less likely to depend specifically on the folding topology. In subsequent NMR experiments, differences in binding modes to the two G4 structures are also demonstrated (Section 2.3.4).

2.3.3 Circular dichroism titrations

2.3.3.1 Introduction to circular dichroism

In the previous section, absorbance spectroscopy was introduced as a way to examine G4 ligand binding, since the interaction of molecules perturbs the electronic structures and leads to observable changes in the respective spectra. This technique is insensitive to the chirality of the absorbing species, since unpolarised light does not itself have a chiral sense. However,

using a combination of a polarising filter and quarter-wave plate, circularly-polarised light may be produced in which the magnitude of the electromagnetic field remains constant but rotates in in a plane perpendicular to the direction of propagation (Figure 2.14a).

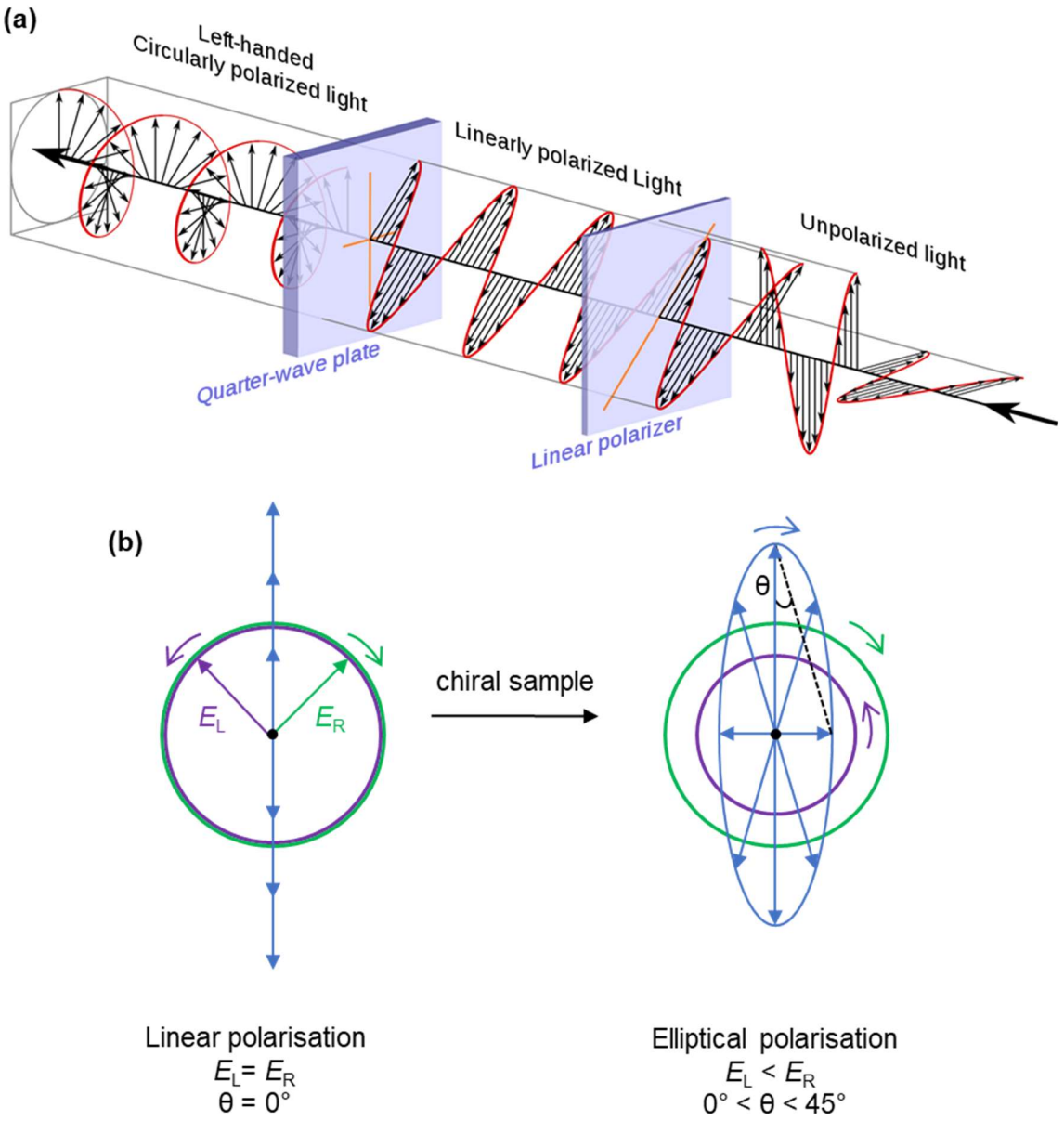


Figure 2.14. Fundamentals of circular dichroism. (a) generation of circularly-polarised light from unpolarised light *via* linearly polarised light; (b) definition of ellipticity based on the interaction of linearly polarised light with a chiral sample. *Figure (a) by Thomas W Cronin, distributed under a CC BY-SA 4.0 license.*

The helical sense of the rotation can either be left- or right-handed, and the two helical forms interact differently with chiral matter due to stereospecific coupling of the transition dipole moments with the rotational sense of the radiation (discussed further below).[191]

Linearly polarised light can be considered a superposition of left- and right-handed circularly polarised light with equivalent amplitude of electric field ($E_L = E_R$), resulting in oscillation of the field vector in a plane perpendicular to the direction of propagation (Figures 2.14a and 2.14b). Following transmission of linearly polarised light through a chiral sample, differential absorption of the left- and right-handed components means that E_L and E_R are no longer equivalent and the superposition of the two components results in an elliptical polarisation, defined by the angle θ (Figure 2.14b). The difference in absorbance of the left- and right-handed components ($\Delta A = A_L - A_R$) commonly measured by spectropolarimeters is related directly to this angle of ellipticity by Eqn. (2.7), which is derived based on the basis of geometrical and physical arguments.[192]

$$\theta = 32.98 \cdot \Delta A \tag{2.7}$$

The observed ellipticity can be expressed as the molar ellipticity of the sample through Eqn. (2.8) to yield an intrinsic value independent of concentration and path length.

$$[\theta] = \frac{100 \cdot \theta}{c \cdot l} \tag{2.8}$$

Circular dichroism in an invaluable technique for the study of G-quadruplex structures, since different folded topologies gives rise to diagnostic spectral signatures.[111,193–195] The physical origin of circular dichroism in G4 is complex, but in broad terms arises from the coupling of the transition dipole moments (exciton coupling) of the stacked guanine residues (Figure 2.15).

The chiral sense of this coupling is determined by the relative orientation of neighbouring bases in the helical arrangement, in turn influenced by the conformations of the glycosidic bond angles (*syn*/*anti*) of the two nucleotides and hence the overall folding topology. Figure 2.15a shows and example of the displacement of the transition dipole moments of two stacked guanine residues which arises when both are in the *anti* conformation. This exciton coupling results in non-degenerate states with opposite rotational strength (R), arising from the combined effect of the electronic and magnetic transition moments which results in a helical charge redistribution (Figure 2.15b). These states are therefore populated differently by left- and right-handed circularly polarised light, resulting in the observed spectral effect (Figure 2.15c).

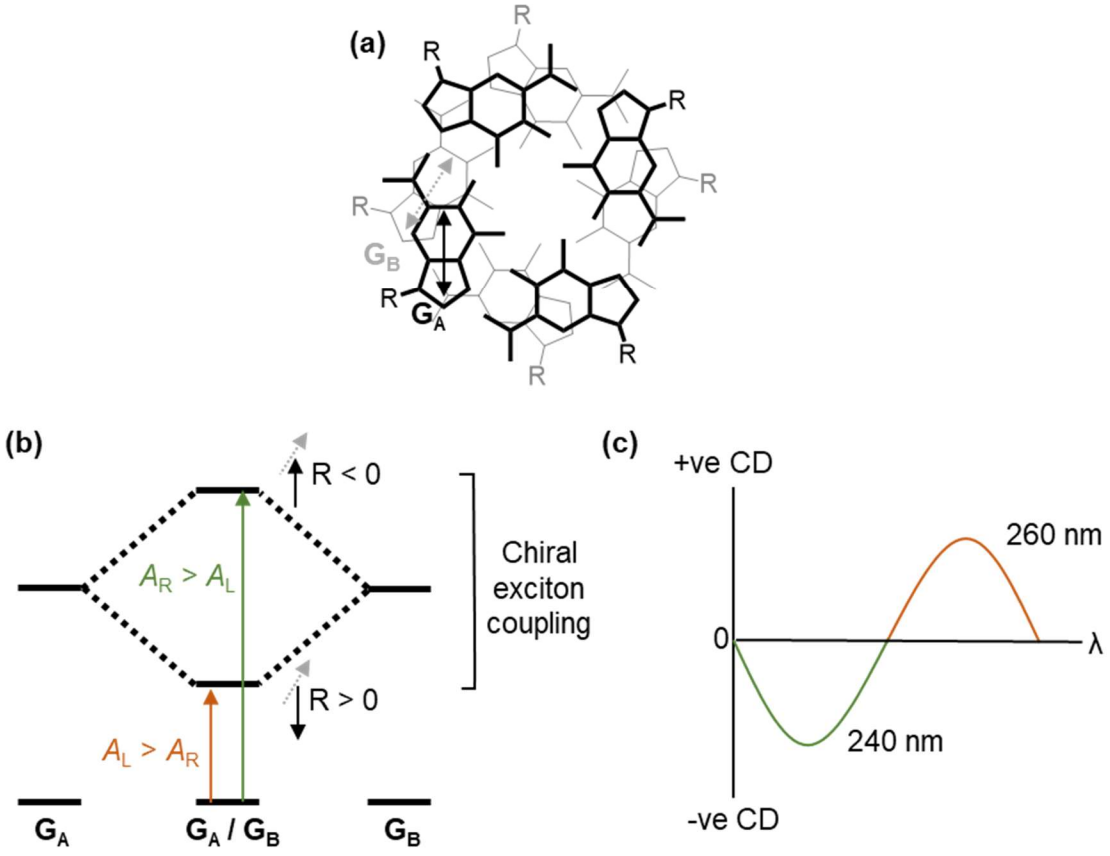


Figure 2.15. Origin of circular dichroism in G-quadruplexes. (a) Simplified representation of two stacked G-tetrads showing the helical displacement of the transition dipole moments for two guanine residues G_A and G_B; (b) exciton coupling of the two transitions leads to non-degenerate states with opposite rotational strength (R) resulting in differential absorption of left- and right-handed circularly polarised light; (c) the resulting CD spectral signature of the exciton coupling represented in (b).

The combination of many such effects over all the guanine-guanine base-steps in the G4 structure gives rise to the characteristic CD signatures of each G4 topology.[195]

CD may therefore be used to examine the effects of ligands on G4 folding, by monitoring the effect of increasing ligand concentration on the characteristic CD bands. Hyperchromic shifts in the bands indicate stabilisation of the folded structure by the ligand,[106] whilst disappearance of spectral features and emergence of new dichroic features can indicate structural rearrangement, such as disruption of the G4 fold or re-folding to an alternative topology.[144,196] Furthermore, since achiral ligands enter a chiral environment upon binding to the G4, CD-

activity may be induced in the ligand-only region of the spectrum which may indicate particular modes of binding.[106]

*2.3.3.2 CD investigations of binding of E-**41** and Z-**41** to G4 structures*

Given the different effects of *E*-**41** and *Z*-**41** observed in the FRET melting assay and confirmed by UV/visible titration studies, circular dichroism titrations were conducted to examine the possible effects of the ligands on different G-quadruplex topologies in physiologically-relevant buffers. Three G4 models were selected: the hybrid (telo23-K^+)[37] and antiparallel (telo22-Na^+)[27] G4 models were used as in the UV/visible titrations. The promoter region of the *c-myc* oncogene (Pu27) was also selected as a model of a parallel G4.[197]

In the titration experiments, the spectrum of the G4 was first recorded in the absence of ligand and subsequently in the presence of increasing stoichiometric equivalents of ligand up to a maximum of 10 molar equivalents (Figure 2.16). The spectra obtained in the absence of ligand (black traces) display the diagnostic features expected for respective G4 sequences. The hybrid G4 (telo23-K^+) is characterised by a positive band at 290 nm with a shoulder at 260 nm and a negative band at 240 nm (Figures 2.16a and 2.16b). Meanwhile the antiparallel form (telo22-Na^+) displays a positive band at 290 nm and a negative band a 260 nm (Figures 2.16c and 2.16d). The parallel form (Pu27) features a strong positive band at 260 nm and negative feature at 240 nm (Figures 2.16e and 2.16f).[111] Only very weak perturbations of the G4 spectra are induced upon the addition of ligand *E*-**41** and *Z*-**41**. The perturbations of the ellipticity do appear to be slightly larger for ligand *E*-**41**, consistent with the observations of the FRET assays and UV/visible titration studies, which indicate this to be the active form of the ligand. For example, in the titration with Pu27, the parallel G4, a hypochromicity in the positive band at 260 nm is observed for ligand *E*-**41** (Figure 2.16e) whilst CD spectrum of the same G4 is comparatively unperturbed by ligand *Z*-**41** (Figure 2.16f). However, overall, the results indicate that ligand binding to G4 occurs without perturbation of the native folding topology, since all the main spectral features of each G4 topology are preserved upon increasing the ligand concentration. Therefore, whilst the difference in G4 binding activity between the two ligand forms was still considered worthy of further investigation towards photopharmacological applications, it was concluded that the *E*-**41** and *Z*-**41** couple did not show strong potential as a means for the ligand-mediated photocontrol of G4 topology for the development of nanodevices, since the ligand is not able to influence the folding dynamics of the DNA structure necessary for such applications.

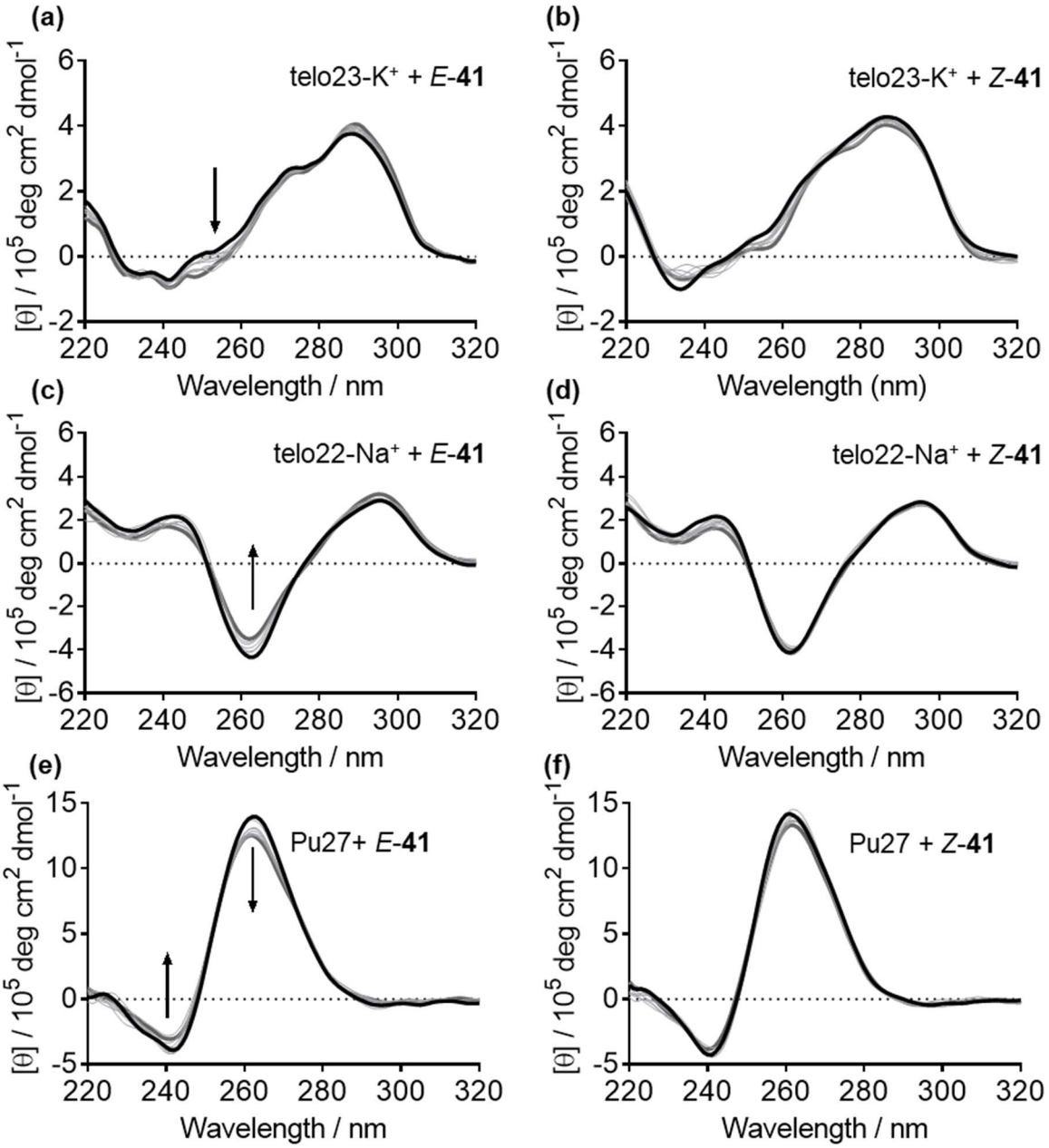


Figure 2.16: CD titrations of (a)-(b) telo23-K^+ hybrid G4, (c)-(d) telo22-Na^+ antiparallel G4 and (e)-(f) Pu27 parallel G4 with (a),(c),(e) ***E*-41** and (b),(d),(f) ***Z*-41.** Black trace: DNA CD spectra in absence of ligand; dark grey trace: CD spectra in the presence of 10 equiv. ligand. Intermediate titration points are shown as pale grey traces. [θ] = molar ellipticity. DNA concentration: 4.2 μM.

2.3.4 NMR studies

Given the inconclusive results of the CD studies regarding the binding modes of the ligands, further investigation of the G4/ligand interactions by ^{1}H NMR spectroscopy was warranted in order to help rationalise the results of the UV/visible titration studies and FRET assays. As well as enabling DNA/ligand complex formation to be more directly observed, these experiments can provide indications of the binding mode of the ligands to the G4 structures.[109,198] In particular, perturbations of the chemical shifts of specific guanine residues are indicative of G-tetrad stacking interactions, whilst line broadening of Hoogsteen-bonded NH (imino) proton resonances without such perturbations suggests binding to regions elsewhere on the quadruplex, such as in the loops or grooves.[106,109]

2.3.4.1 Studies with hybrid (telo23-K^{+}) G4

First the interaction of *E*-**41** and *Z*-**41** with hybrid (telo23-K^{+}) G4 was investigated, since the FRET and UV/visible titration studies indicated *E*-**41** selectively associates with this form of the telomeric G4 (as opposed to the antiparallel G4 formed in sodium conditions) whilst ligand *Z*-**41** remained comparatively ineffective. The NMR spectra of this G4 was reported by Phan *et al.* in 2006 and the sequence was found to exist as a mixture of major (~ 70%) and minor (~ 30%) conformations under the NMR conditions.[37] This makes it rather difficult to monitor chemical shifts in the aromatic region of the spectrum as many resonances are overlapped, but the imino resonances were confidently assigned in accordance with the reported data.

Upon titration with ligand *E*-**41**, significant changes to the telo23 imino resonances were observed (Figure 2.17a). At 1:1 G4/ligand stoichiometry, whilst several resonances remain distinct (most clearly G10, G9, and G21, indicated by numbered resonances), other resonances appear almost absent (especially G3, G15 and G23, indicated with asterisks) and at 2:1 stoichiometry appear to have disappeared entirely. The apparent loss of imino resonances could result from a number of factors. One possibility is precipitation of the complex at the high concentrations (~ 200 μM) required for NMR experiments, but since the intensity of several imino resonances are preserved throughout the titration this possibility can likely be discounted. Another possible explanation is that intermediate exchange between free and bound ligand states on the NMR timescale (resulting in the case of strong binding ligands which exhibit slow dissociation kinetics from the biomolecule) results in line broadening that ultimately leads to the apparent disappearance of the signals (Figure 2.18 and caption).

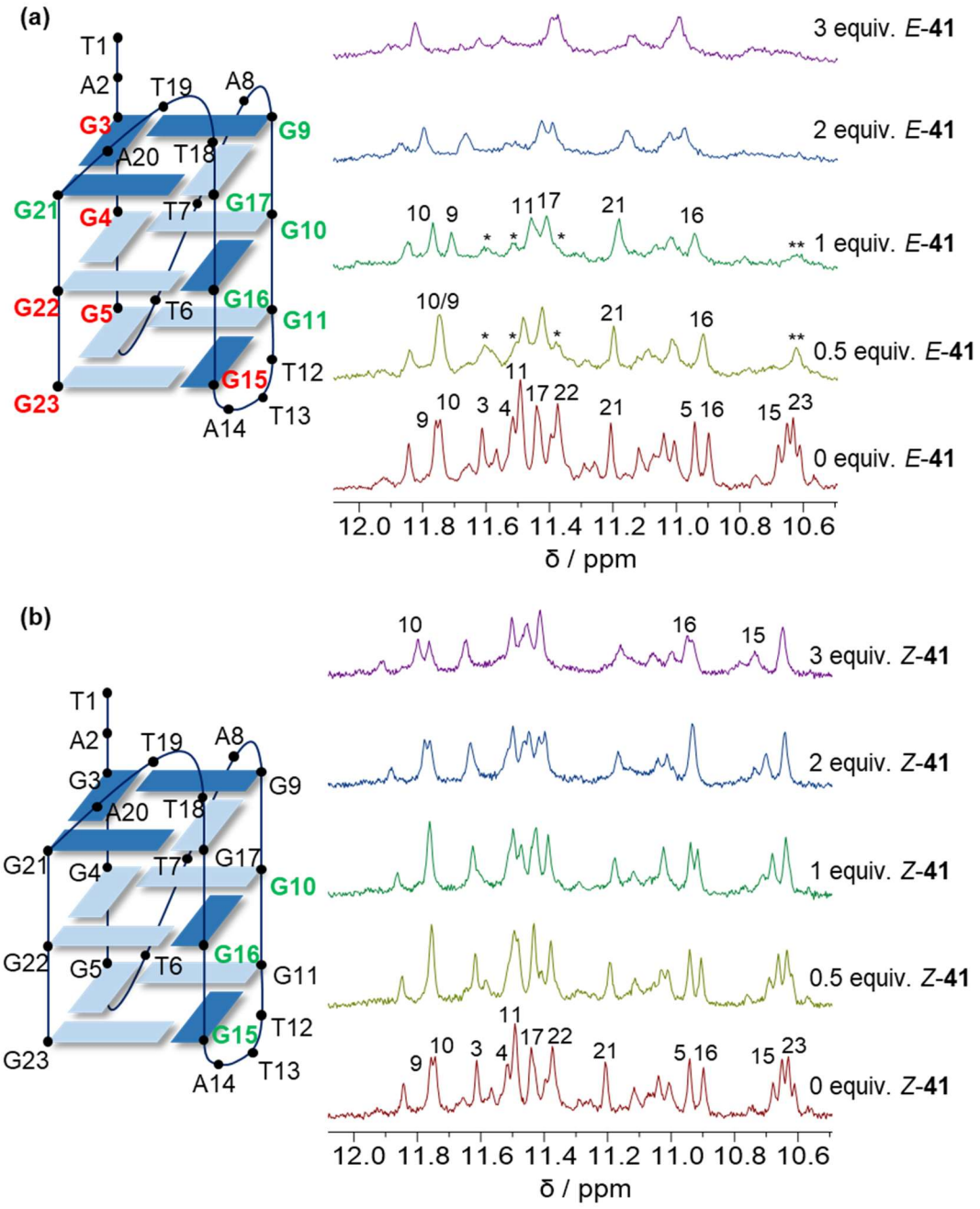


Figure 2.17: ^{1}H NMR titration studies (Hoogsteen NH imino region) of hybrid G4 (telo23-K^{+}) titrated with (a) *E*-**41** and (b) *Z*-**41**. The cartoon representation in each case labels perturbed imino resonances in green and disappearing resonances in red. DNA concentration: 185 μM.

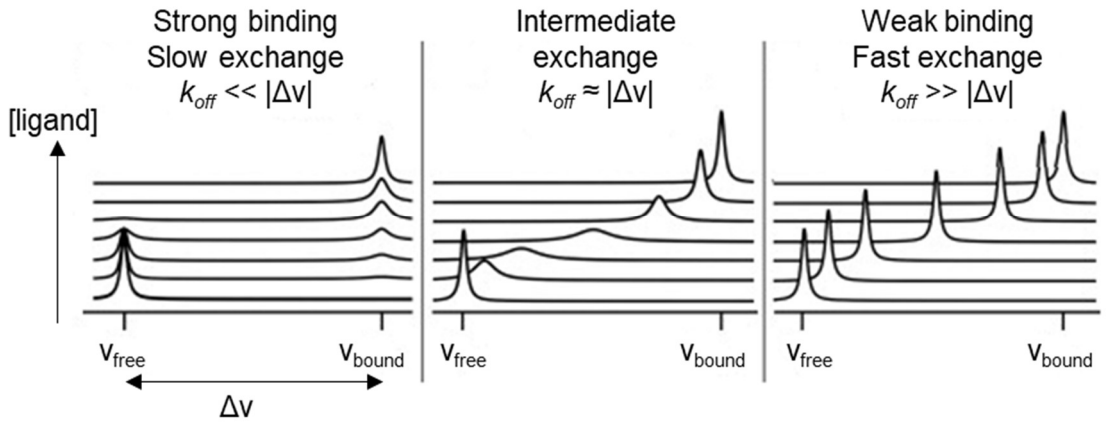


Figure 2.18: Possible exchange broadening effects in NMR titrations. A strongly bound ligand exhibits slow dissociation kinetics from the DNA on the NMR timescale, meaning the k_{off} rate is slower than the difference in frequency (Δv) between the resonances corresponding to the free DNA (v_{free}) and bound complex (v_{bound}). The free and bound DNA states are therefore distinct on the NMR timescale, and both resonances are observed with intensity proportional to the mole fraction of each species. In the fast exchange regime, exhibited by weakly bound ligands with fast dissociation kinetics, the free and bound states rapidly interconvert on the NMR timescale and the position of the chemical shift appears as weighted average of those of the free and bound species. In the intermediate regime the free and bound complexes interconvert on the NMR timescale, resulting in line broadening accompanied by a change in chemical shift. *Figure adapted with permission © Elsevier.*[199]

In this case, it should be possible to observe the recovery and sharpening of signals corresponding to the bound complex as the concentration of bound complex increases (Figure 2.18). That no such recovery is observed at high ligand concentration suggests that that the disappearance of imino signals may not be the result of exchange broadening. The remaining possibility is that *E*-**41** interacts with hybrid G4 through an intercalative mechanism, which requires partial disruption of the G-tetrad network leading to the disappearance of certain resonances. It should be noted that the circular dichroism experiments did not indicate disruption of the folded structure (Figure 2.16a). However, the concentrations required for the NMR experiments are approximately 50-fold higher than those employed in the CD studies.
Meanwhile, the comparatively small effects on the telo23 imino resonances observed upon titration with *Z*-**41** (Figure 2.17b) are consistent with the previous results that indicate this ligand interacts less strongly with the G4 than the *E*-**41** counterpart. Indeed, all imino resonances remain distinct throughout the titration. Instead, only minor chemical shift

perturbations can be observed for G10, G16 and G15 resonances, indicating weak interaction with this specific region of the G4. The NMR results therefore support the results of the other biophysical assays and confirm that the stilbene configuration is an important determinant of ligand binding mode and therefore for the affinity observed for the hybrid G4.

2.3.4.2 Studies with antiparallel (telo22-Na^+) G4

The binding of the *E*-**41** and *Z*-**41** ligands to antiparallel telo22-Na^+ G4 was also examined. Though the ligand affinity for this G4 sequence observed in the FRET and UV/visible titration assays appeared to be rather low, the concentrations required for NMR experiments are much higher (~180 μM versus ~10 μM) meaning complexation can be observed even for comparatively weak-affinity ligands. Moreover, since telo22-Na^+ adopts a single folded conformation in solution, assignment and interpretation of the spectra, including the aromatic and carbohydrate resonances is much more straightforward.[27]

The binding of *E*-**41** to the antiparallel G4 was first examined (Figure 2.19). In comparison to the effect of this ligand on the hybrid telo23 G4, only minor changes in the imino region of the spectrum were detected. Rather than the disappearance of specific resonances, a gradual change in chemical shift of the imino resonances is observed throughout the titration, indicative of fast exchange. Specific perturbations of the G8 and G4 resonances were observed (Figures 2.19a and 2.19c) with the other imino resonances only comparatively weakly affected. Owing to the comparative simplicity of the telo22 spectra, many of the aromatic signals could be resolved in 2D NOESY data using the assignments reported by Wang and Patel.[27] This reveals specific chemical shifts in the aromatic protons of the two lateral loops, particularly the A19H8 and A7H8 protons, and the H2'/H2'' sugar ring protons of T6 (Figures 2.19b and 2.19c, full spectrum in Appendix A2, Figure A2.14). Meanwhile, no significant chemical shift perturbations are observed for residues corresponding to the middle or lower G-tetrads, nor the diagonal loop (Figure 2.19d). Taken together, these data indicate the ligand specifically targets the top face of telo22, perhaps stacking with the G-tetrad owing to its planar structure.

Meanwhile, weaker chemical shift perturbations are observed for the same G4 sequence on titration with ligand *Z*-**41** (Figure 2.20 and Appendix A2, Figure A2.15). Though the imino resonances experience weak perturbation (Figure 2.20a), the effect is more subtle than for the *E*-**41** counterpart. Furthermore, the specific perturbation of the A7 and A19 residues of the lateral loops is not observed (Figure 2.20b). Again, these results demonstrate the importance of ligand planarity in activating the specific binding mode of the stiff-stilbene ligand.

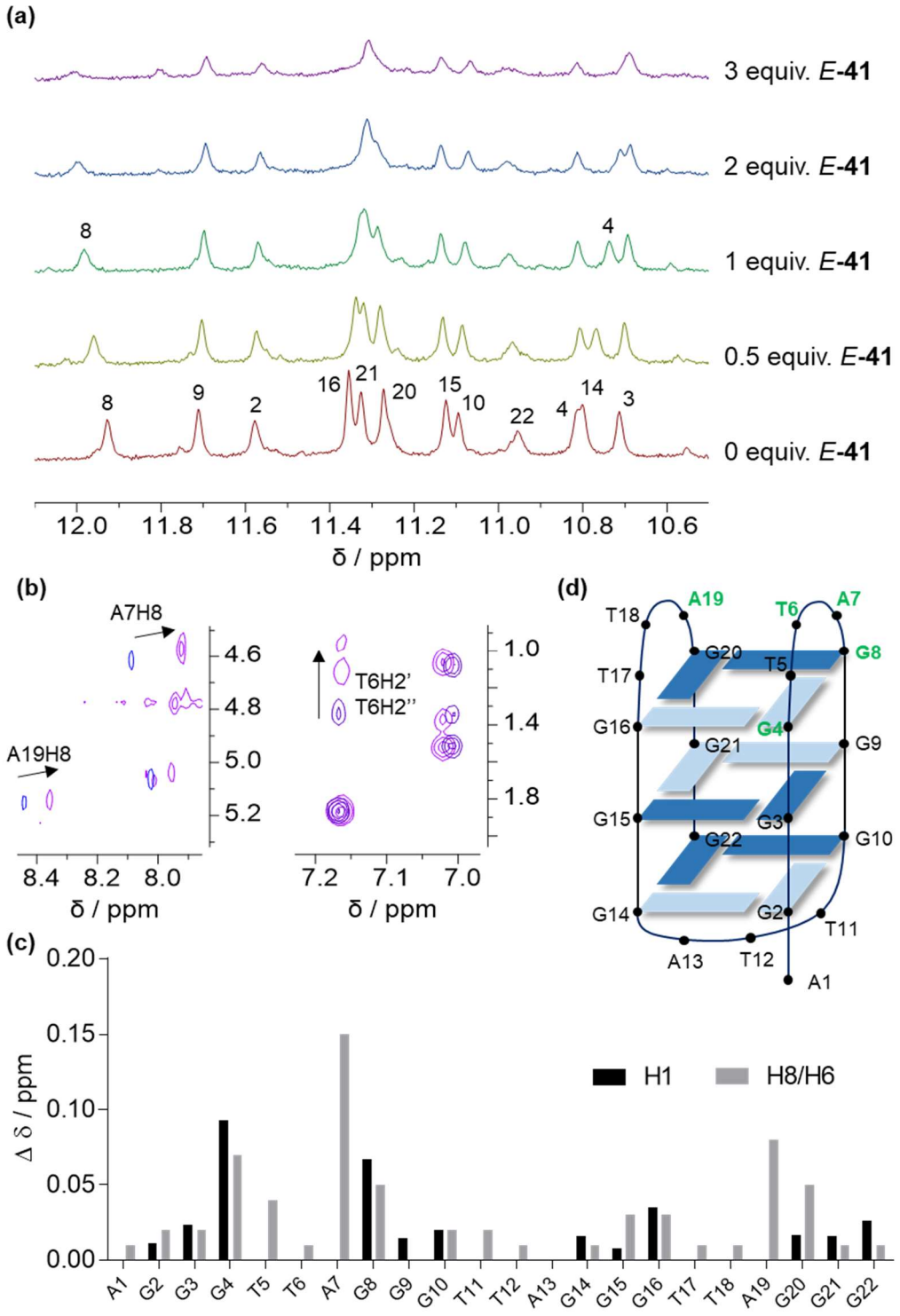


Figure 2.19: ^{1}H NMR titration studies examining the binding of *E*-**41** to antiparallel (telo22-Na^{+}) G4. (a) stacked imino spectra; (b) relevant regions of the 2D NOESY spectra of the free DNA (blue) and with 2 equiv. *E*-**41** (pink); (c) chemical shift perturbations of imino and aromatic resonances of DNA upon addition of 2 equiv. *E*-**41**; (d) cartoon representation of the antiparallel G4 with chemical shift perturbations > 0.05 ppm highlighted in green.

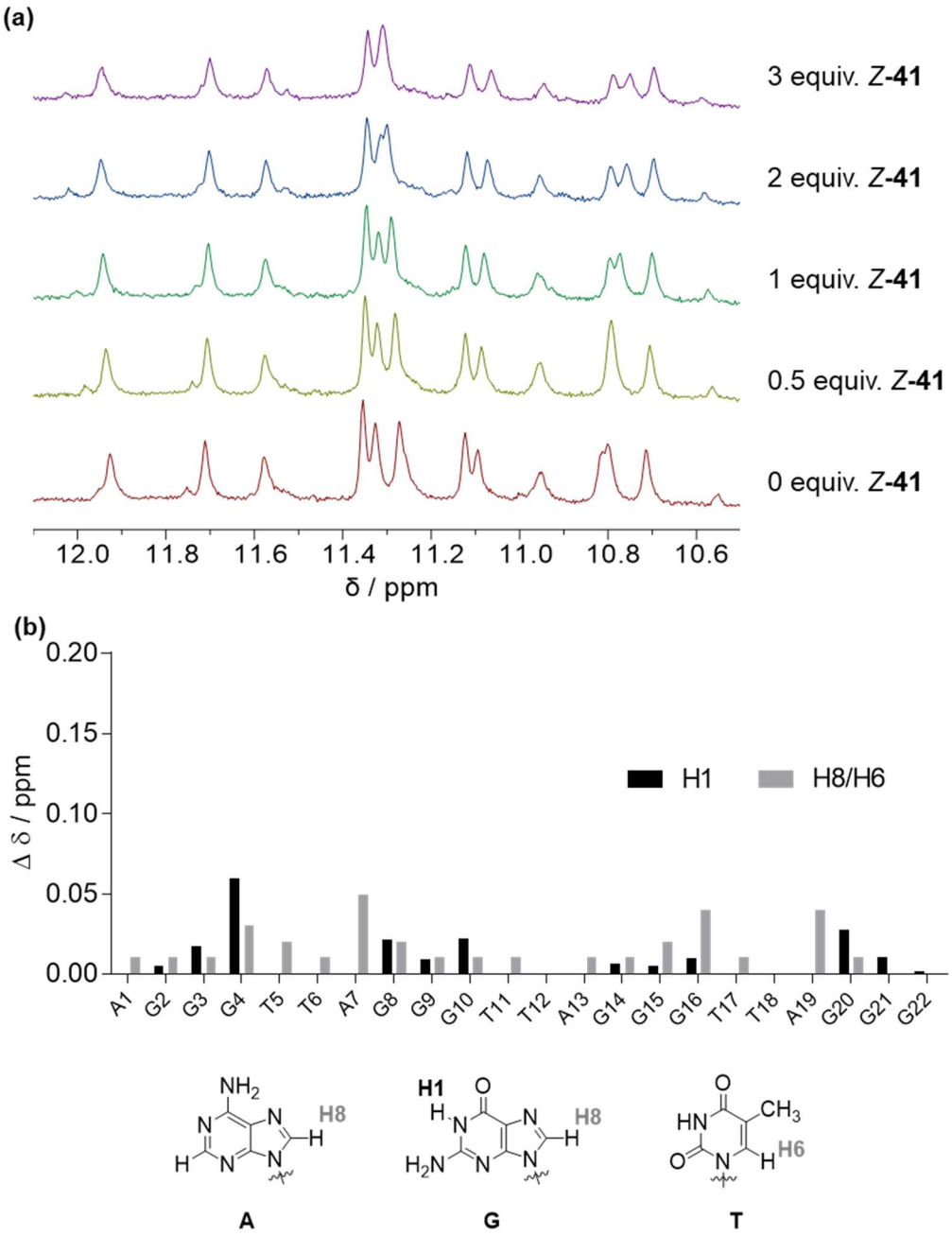


Figure 2.20: NMR titration studies examining the binding of *Z*-**41** to antiparallel (telo22-Na^+) G4. (a) stacked imino spectra; (b) chemical shift perturbations of imino and aromatic resonances of DNA upon addition of 2 equiv. *Z*-**41.**

2.4 *In vitro* assays

The results of the biophysical assays demonstrated that the stiff-stilbene ligands can selectively target G4 in physiologically-relevant conditions. Importantly, the *E*/*Z*-**41** ligand pair

acts as a proof-of-concept that isomerisation of the stilbene core could act as a means to control G4 ligand affinity. However, despite differences in activity of the ligand pair, the affinity of the active *E*-**41** ligand for the preferred hybrid G4 ($K_a \approx 10^4$) is not particularly high. This raised the question of whether the ligands developed in this chapter were sufficiently potent DNA binding molecules for the promising activities observed in the binding studies to translate into corresponding activities in biological systems. Before embarking upon efforts to study the photochemistry of *E/Z*-**41**, it was considered more worthwhile to evaluate the toxicity of these ligands in *in vitro* cellular and parasitic models in order to establish the potential of the *E/Z*-**41** pair as a photopharmacological strategy, or whether further optimisation of the G4-binding properties was first necessary to progress towards this goal.

Ligands *E*-**41** and *Z*-**41** were therefore screened for activity against two mammalian cell lines: HeLa (a cervical cancer cell line) was chosen as a tumoral model whilst MRC-5 (lung fibroblasts) was selected as a non-cancerous model. Meanwhile, to evaluate the potential of the compounds as antiparasitic agents, toxic activity was measured in the protozoan parasites *T. brucei* and *L. major*.[52]

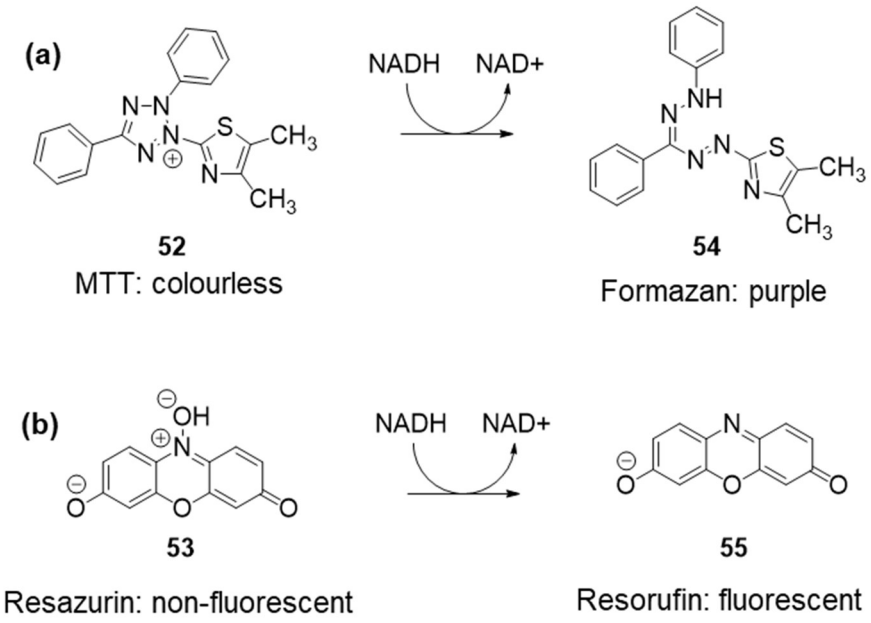


Figure 2.21. Chemical basis of (a) MTT and (b) alamarBlue cell viability assays. The products generated by the reducing environments of normally functioning cells are detected by absorbance (formazan) or fluorescence (resorufin) and the readout is proportional to the quantity of viable cells.

2.4.1 Basis of the viability assays

The MTT and resazurin (alamarBlue) assays use redox processes to measure cellular viability. Live cells maintain a reducing environment owing to the production of reducing agents (such as NADH) during metabolism.[200] Both MTT (**52**) and resazurin (**53**) are susceptible to reduction by these species, generating products (formazan **54** and resorufin **55**) which may be detected through fluorescence or absorbance spectroscopy (Figure 2.21).[201,202] The respective readouts are therefore proportional to the quantity of viable cells under the experimental conditions. To measure compound toxicity, cells and parasites were incubated with increasing concentrations of compound for 72 hours and the viability under each set of conditions was measured using alamarBlue (mammalian cells and *T. brucei*) or MTT (*L. major*). Normalisation against a negative control (cells not treated with compound) allows the toxic effect of the compound to be quantified as the concentration which results a 50% reduction in viability relative to the control, known as the EC_{50} (effective concentration).

2.4.2 Toxicity studies in cells and parasites

First the activity of the pair or ligands *E*-**41** and *Z*-**41** was inspected against a small panel of mammalian cells and parasites (Figure 2.22 and Table 2.3). Though a 5-fold difference in EC_{50} between the *E*-**41** (EC_{50} = 46 ± 8 μM) and *Z*-**41** (EC_{50} = 9 ± 3 μM) isomers is observed against HeLa cells, it is the *Z* isomer which is more efficacious. In light of the biophysical assays, which demonstrate this to be the poorer G4 ligand, this suggests that the mechanism of toxicity is unlikely to be related to G4 binding for these two compounds. Furthermore, the toxicity of both ligands to the non-tumoral cell line is similar to that observed in the cancerous model (e.g. for *Z*-**41**, EC_{50} (HeLa) = 9 ± 3 μM, EC_{50} (MRC5) = 9 ± 1 μM). Since more potent G4 ligands have previously been demonstrated to selectively target tumoral cells through a variety of proposed mechanisms (Section 1.3.1), the similarity in toxicity between the different cell lines tentatively suggest alternative mechanisms of action that are not G4-dependent. Therefore, it appears that the G4-binding activity of the *E/Z*-**41** couple is not sufficiently strong for the trends observed in the biophysical assays to translate into *in vitro* performance.

Turning attention to the activity of the compounds against the parasitic strains, significantly higher toxicity against *T. brucei* is observed, with EC_{50} values recorded in the nanomolar range (e.g. EC_{50} *Z*-**41** = 12 ± 5 nM). Whilst promising, these do not correlate at all with the G4-binding affinity of the molecules.

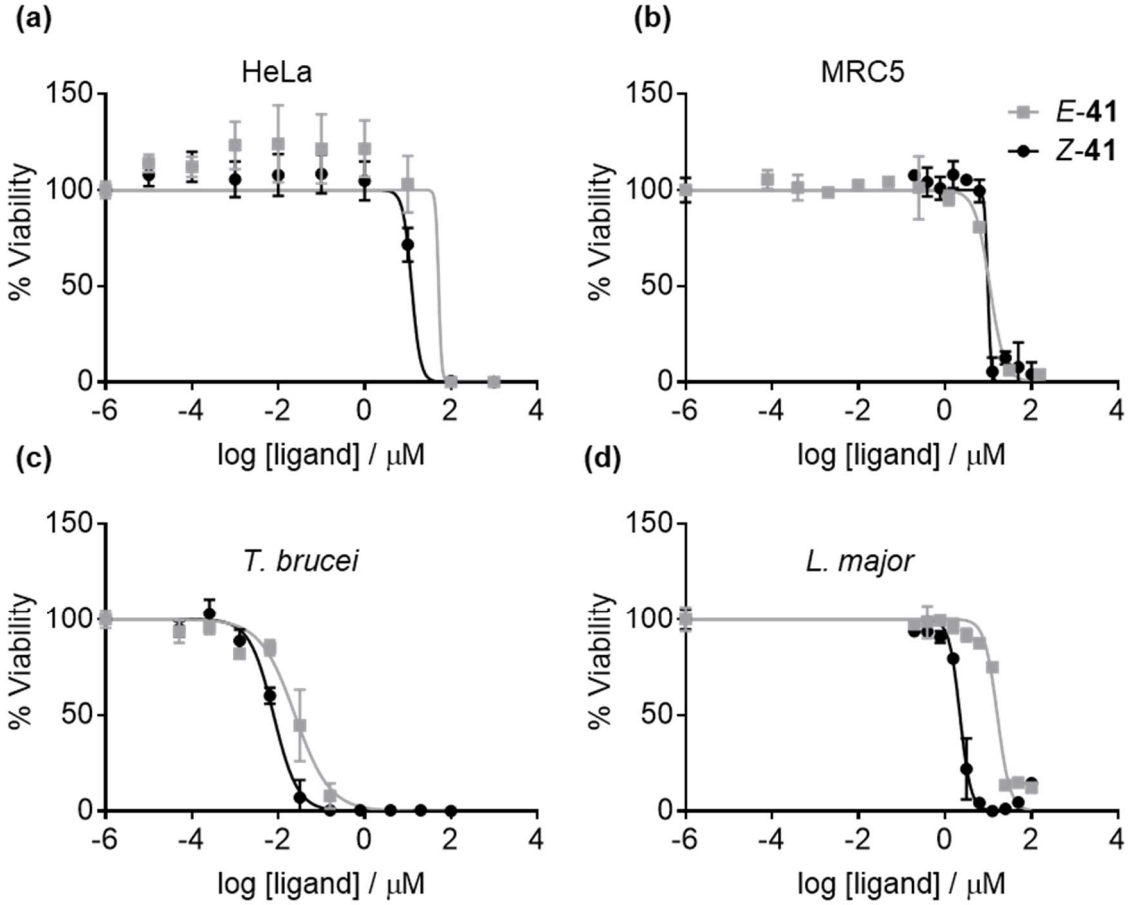


Figure 2.22: Dose-response curves for ligands *E*-**41** and *Z*-**41** against (a) HeLa and (b) MRC5 cells, and (c) *T. brucei* and (d) *L. major* parasites. Exposure time: 72 h.

Table 2.3: EC_{50} values for compounds *E*-**41** and *Z*-**41** against the cellular and parasitic models studied. Selectivity index = EC_{50} (MRC5) / EC_{50} (HeLa/*T. brucei*/*L. major*)

	EC_{50} / μM				Selectivity index		
Ligand	HeLa	MRC5	*T.brucei*	*L.major*	HeLa	*T.brucei*	*L.major*
E-**41**	46 ± 8	10 ±1	0.036 ± 0.001	16 ± 0.4	< 1	280	< 1
Z-**41**	9.0 ± 3	9.0 ±1	0.012 ± 0.005	2.3 ± 0.07	1.0	750	3.9

The *Z*-**41** isomer is the more active form against both parasitic strains (between three- and seven- fold decrease in EC_{50} compared to *E*-**41**), again suggesting that the biological activity is not a result of a G4-mediated pathway. Nonetheless, the compounds are both significantly more toxic to *T. brucei* than non-tumoral cells (up to 750-fold in the case of *Z*-**41**) and further

investigations into the mechanism of the activity of these compounds would be worthwhile to explore their potential as a new class of potential anti-trypasonomal agents. The compounds are less efficacious against the *L. major* organism, but nonetheless *Z*-**41** displays a 4-fold selectivity against the non-tumoral model.

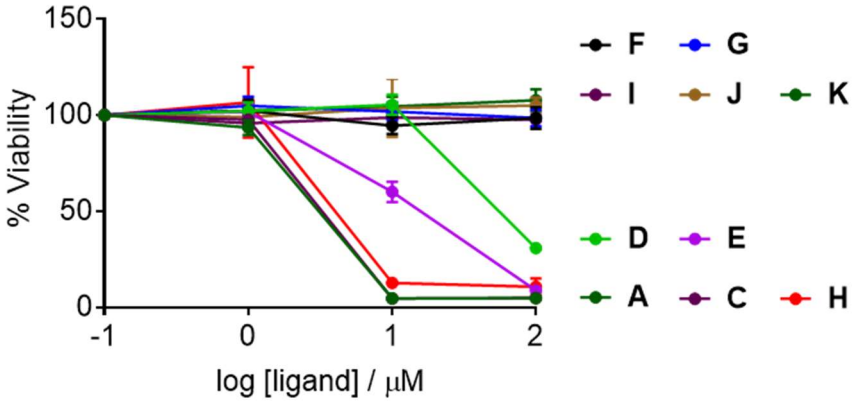


Figure 2.23: Toxicity screen of the compound library **A** and **C-H** against HeLa cells. Exposure time: 72 h.

Finally, the remainder of the compound library (compounds **A** and **C**-**H**) was screened for toxicity against HeLa cells to further examine the influence of the ligand structure on biological activity (Figure 2.23). Com pounds **A** and **C**, direct analogues of *E*-**41** only differing in the length of the alkyl spacer chain, displayed similar activities ($EC_{50} \approx 50$ μM). The only other toxic compounds appeared to be **D**, **E** and **H** and therefore no correlation with the G4 binding properties observed in the FRET assay was found. For example, **E** and **F** differ only in length of the spacer unit, which does not much impact the thermal stabilisation of G4 in the FRET assay but appears to significantly impact the toxicity. Again, these results highlight that the biologically activity of the compound series developed in this chapter is unlikely to result from G4 binding.

2.5 Chapter conclusions

The work described in this chapter details the initial efforts of this research project towards the development of a new series of photoresponsive G4 ligands with the overall aim of achieving a G4-targeting ligand that can be regulated in physiological conditions towards the development of G4-targeted photopharmacological agents. The limitations in the current state-of-the-art (azobenzene derivatives) have been critiqued, leading to the design of a new family of ligands based on the stiff-stilbene photoswitch. A model photoswitchable ligand pair (*E*-**41** and *Z*-**41**) was conceived and synthesised, and the binding of the two isomeric derivatives

was investigated using a suite of spectroscopic methods. This demonstrated that the stiff-stilbene scaffold is a promising fragment for the design of G4 ligands. Indeed, this first attempt to engineer a G4 ligand derived from this scaffold yielded an active compound (*E*-**41**) with affinity at the micromolar level and selectivity against duplex DNA. Critically, the *Z*-**41** isomer was shown to be inactive as a G4 ligand, indicating that the G4 binding activity of this compound series could conceivably be controlled by photoisomerisation. NMR spectroscopy confirmed differences in binding modes between the two ligands. Meanwhile, investigations into the structure-activity relationships were undertaken (compounds **A**-**H**), though unfortunately these did not return any significant increase in G4-stabilising ability. Unfortunately, the biological activity of the isomeric forms, whilst of interest (particularly the potency of the compounds against parasitic organisms) did not match the ability of the ligands to bind G4. It was concluded that improvements to the G4-recognition properties in this compound series were needed in order to build more convincing evidence that stiff-stilbene derivatives may eventually target G4 *in vivo*.

3 *N*-methylpyridinium-decorated stiff-stilbene G4 ligands

This chapter is based upon work published in *Angewandte Chemie International Edition*[203] and *Chemistry: A European Journal*[169] of which I am first author. I was responsible for all elements of ligand design, synthesis and biophysical studies, which were undertaken at the University of Bristol, UK. The DFT calculations were performed by Miss Marta Duchi in Dr. Thomas A.A. Oliver's group at the University of Bristol. I performed the biological assays at the Institute of Parasitology and Biomedicine (CSIC), Granada, Spain, with assistance from Dr Pablo Peñalver. Dr Peñalver also collected the microscopy images. The molecular dynamics studies were performed by Dr Susanta Haldar in Professor Adrian J. Mulholland's group at the University of Bristol.

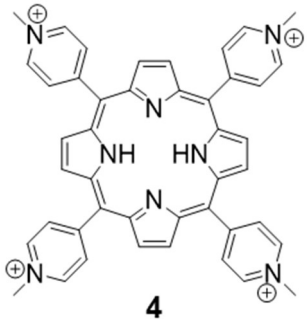
3.1 Introduction

The previous chapter described a proof-of-concept study of the design of G4 ligands derived from stiff-stilbene with the goal of achieving photocontrol of ligand activity in physiological conditions. Promising activities were achieved: the first-generation compounds *E*-**41** and *Z*-**41** selectively bound G4 in preference to duplex, and the affinity of the ligand for G4 was shown to depend on the geometric configuration of the stilbene moiety, suggesting photocontrol of the ligand activity could ultimately be achieved. However, the relatively modest affinity of the active molecule for G4 did not appear sufficient for G4-mediated toxic effects in mammalian cells and parasites to be observed *in vitro.* Optimisation of the ligand design was therefore desired, in order to create a more potent class of compound that could compete with high-affinity compounds previously reported. This chapter describes the design of a second generation of stiff-stilbene ligands bearing *N*-methylpyridinium cationic moieties in contrast to the flexible amine-terminated side-chains featured in the prototype *E/Z*-**41** ligands. The rationale for the compound design is first explained and the synthesis of a set of analogues is briefly described. The results of the biophysical assays are then discussed which reveal this class of compounds display markedly improved G4 binding properties and also dramatic effects on G4 folding topologies. As well as the suite of biophysical techniques described in the previous chapter, results of binding simulations from molecular dynamics and metadynamics studies (conducted by Dr Susanta Haldar) are presented in light of NMR data gathered by the author. Finally, the toxicity of the new compounds in cellular and parasitic models is reported, which suggests that the improved G4-targeting properties of the new ligands translate into promising selective effects against disease models and indicate a role for stiff-stilbene ligands in photopharmacology.

3.2 Ligand design and synthesis

3.2.1 Design concept

In search of a means to increase the G4-binding ability of the stiff-stilbene scaffold, it was noted that *N*-methylated aromatic moieties are commonly found in G-quadruplex ligands that are reported to display high binding activity (see Section 1.3.1). For example, TMPyP4 (**4**, Figure 3.1), bearing four *N*-methyl pyridinium groups, is known as a potent G4 ligand but also targets duplex DNA.[104] It was thus reasoned that the combination of the *N*-methylpyridinium moiety with the stiff-stilbene core may improve G4 ligand binding potency by extending the ligand surface area and cationic charge, whilst possibly retaining the selectivity of the scaffold observed in the amine-decorated variants (Figure 3.2).



4

Figure 3.1. TMPyP4 (**4**), a potent G4 binding chemotype containing four *N*-methylpyridinium moieties.

3.2.2 Structural considerations

To examine the design hypothesis, the pyridinium ligands *E*-**56** and *Z*-**56** were sought as direct analogues of the prototype amine-decorated *E/Z*-**41** pair (Figure 3.2a). The molecular structures of these ligands were examined by DFT calculations (Figure 3.3) to assess the effect of the new substituent on the ligand geometry. From the computational results, it appears that the presence of the pyridinium moieties confers a distortion from planarity in the *E* isomer (Figure 3.3c), though the molecule nonetheless possesses a large extended surface that could be expected to stack or intercalate the G-tetrad. The increased ligand surface area can be seen by comparison of the dimensions of *E*-**56** (Figure 3.3d) with the unfunctionalised stilbene core (Figure 3.3b), and appear more suited to overlap with the surface of a G-tetrad (Figure 3.3a). Meanwhile, *Z*-**56** adopts a V-shaped conformation, with a ~ 20° bite angle between the two halves (Figure 3.3e). This geometry suggests that only one half of the

molecule can be expected to stack with the planar G-tetrads at any one time, and the structure was therefore hypothesised to be a weaker ligand than *E*-**56**.

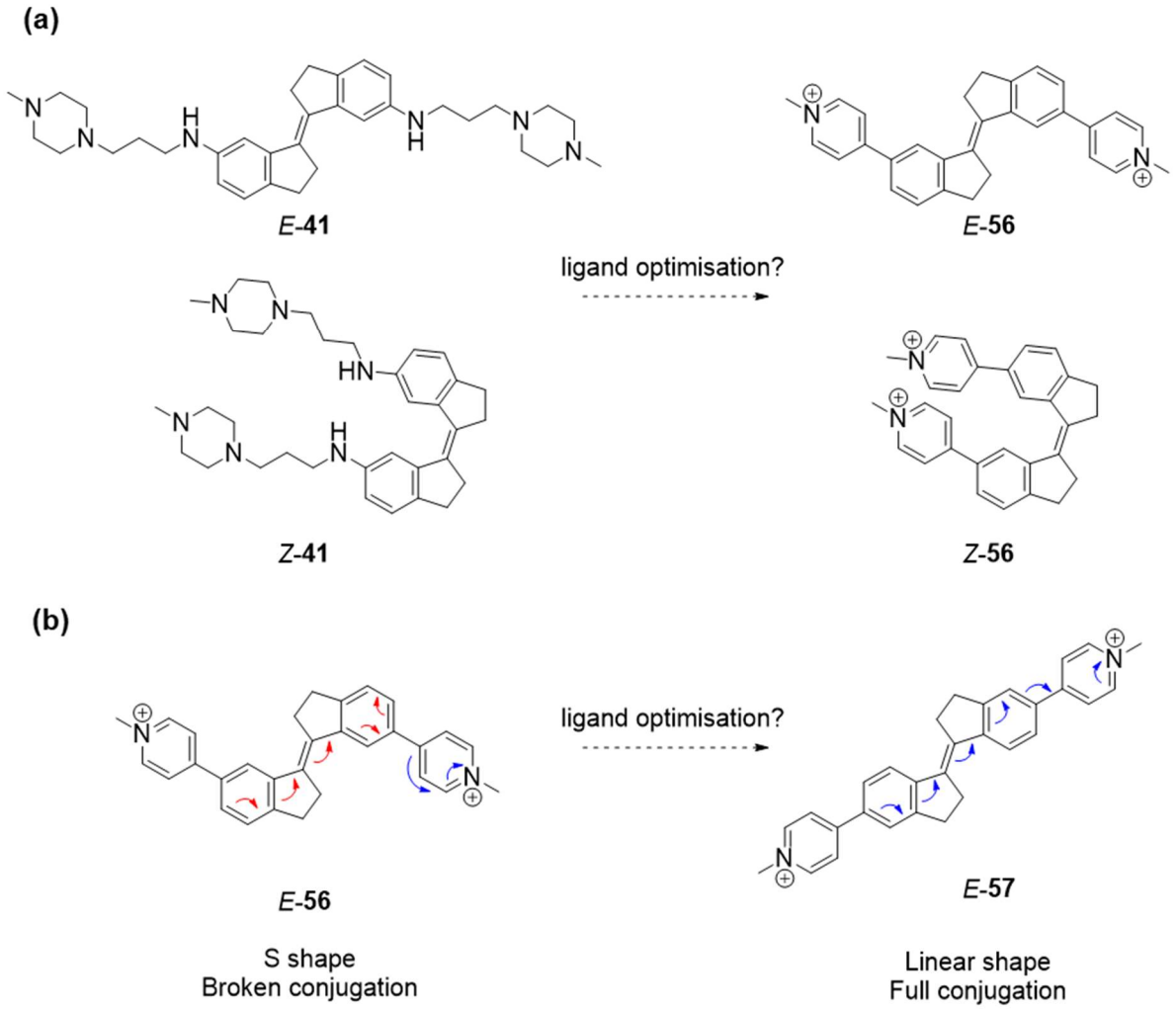


Figure 3.2. Ligand optimisation strategies pursued in this chapter. (a) Replacement of flexible transiently-charged side-chains with rigid cationic *N*-methylpyridinium moieties; (b) effect of substitution pattern on ligand shape and electronics.

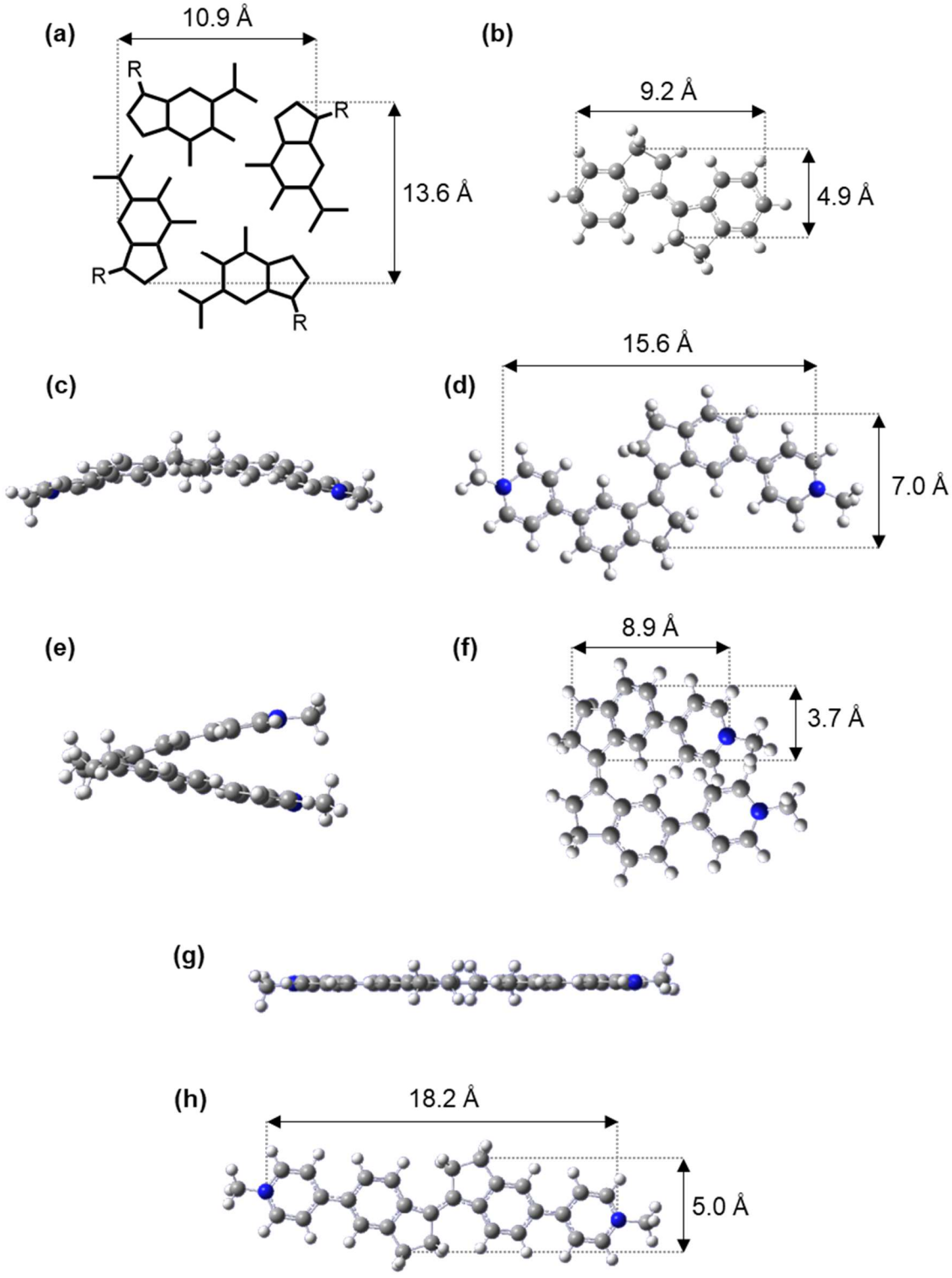


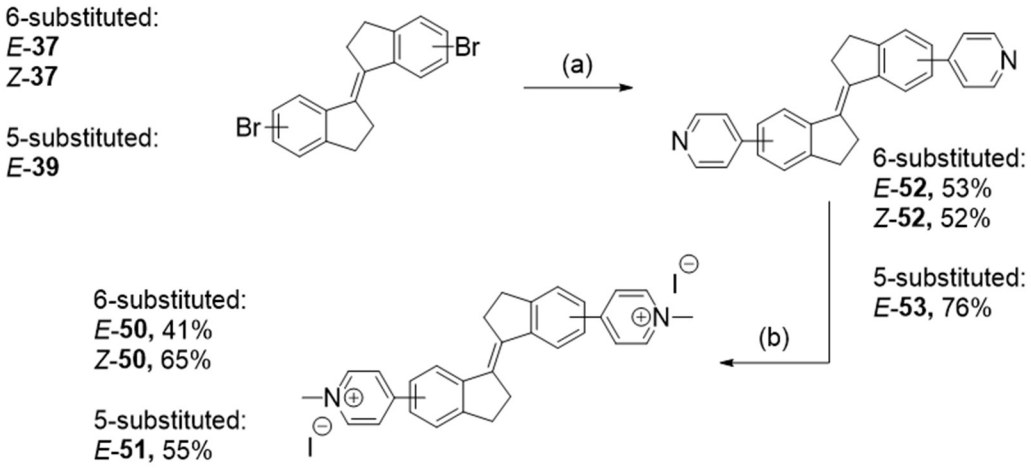
Figure 3.3: Investigation of candidate ligand structures by DFT. (a) Dimensions of a G-tetrad; (b) dimensions of the unsubsituted stiff-stilbene core; (c)-(h) calculated ground-state strucuture of (c)-(d) *E*-**56**, (e)-(f) *Z*-**56** and (g)-(h) *E*-**57** shown (c),(e),(g) side-on and (d),(f),(h) top-down.

In Chapter 2, the position of the amine-derived substituents on the stilbene core was varied (e.g. ligand **D** vs. ligand **E**). This modification appeared to exert little effect on the G4 binding properties as measured in the thermal melting assay, but this is perhaps unsurprising given the conformational freedom of the alkyl spacer units, which may allow the terminal methylpiperazine moieties to adopt a favourable orientation irrespective of the connectivity to the central scaffold. However, the pyridinium ligand variants contain only one rotatable bond (between the stiff-stilbene core and the *N*-heterocycle) and therefore the position of the substituent exerts a much greater influence on the molecular shape of the ligand. In particular, installing the pyridinium moiety *para* to the stilbene olefin (at the 5-position), rather than *meta* (at the 6-position) renders a more rod-like shape to the molecule (ligand *E*-**57**) which could reasonably be expected to modulate the affinity of the ligand to DNA structures (Figure 3.2b). DFT optimisation of the ground state molecular geometry indicates the molecule to be fully planar, unlike *E*-**56** (Figures 3.3g and 3.3h). This difference is proposed to arise from the difference in electronics between the two isomers (Figure 3.2b). In *E*-**56** the positive charge may be delocalised over the two halves of the molecule since the pyridinium moiety is conjugated to the central olefin, and therefore the planar conformation is likely favoured to maximise orbital overlap of the two π-systems. This extended delocalisation is not possible in the 6-substituted isomer *E*-**56**, resulting in a distortion from planarity being favoured on steric grounds.

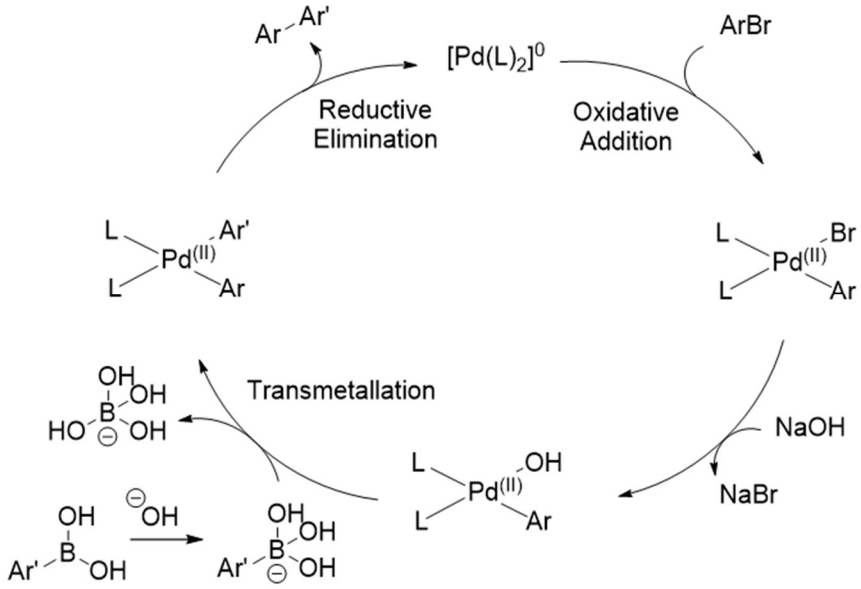
3.2.3 Synthesis of the target compounds

With a prototype set of target compounds identified to test the design hypotheses, efforts turned towards their synthesis. The synthetic route to the pyridinium-substituted compounds was undertaken starting from the same bis-bromides (*E*-**43**, *Z*-**43** and *E*-**45**) previously prepared for the synthesis of the amine-derived analogues in Chapter 2 (Scheme 3.1). In this instance, the pendant pyridine moieties were attached by a double Suzuki coupling to 4-pyridinylboronic acid to furnish the bis-pyridinium compounds (*E*-**58**, *Z*-**58** and *E*-**59**) in moderate to good yields. Like the Buchwald-Hartwig coupling introduced in Chapter 2, the Suzuki reaction is also promoted by a Pd(0) catalytic cycle (Scheme 3.2).[204] In this case, following oxidative addition of the aryl bromide, transmetallation between the Pd(II) species and the boronic acid partner occurs, generating the bis-aryl palladium(II) intermediate which undergoes reductive elimination to form the cross-coupled product and regenerate the active palladium(0) catalyst. The transmetallation is promoted by the presence of base, which generates the active boron-ate complex from the boronic acid.

From the bis-pyridinium compounds, alkylation with iodomethane furnished the desired compounds as bis-iodide salts. Full synthetic details are provided in the experimental section and the ^{1}H and ^{13}C spectra of the final compounds in Appendix A3.



Scheme 3.1. Synthetic route to *N*-methylpyridinium stiff-stilbene ligands *E*-**56,** *Z*-**56** and *E*-**57**. Reagents and conditions: (a) 4-pyridylboronic acid, $Ph(PPh_3)_4$ (20 mol %), Na_2CO_3, PhMe/EtOH (9:1), 100 °C, 16 h; (b) iodomethane, DMF, 90 °C, 1 - 16 h.



Scheme 3.2: General mechanism of the Suzuki reaction

3.3 Biophysical studies

With the compound library in hand, efforts turned to evaluating the potential of the compounds as new G4 binding molecules using a similar panel of techniques as described in Chapter 2.

3.3.1 FRET thermal melting assays

DNA binding activity of the three compounds was initially assessed using the FRET thermal melting assay as previously introduced in Chapter 2 (Section 2.3.1). All ligands were screened against the four G4 models previously described, namely the human telomeric sequence (in potassium- and sodium- containing buffer) the *c-myc* oncogene model and the EBR1 G4 from the *T. brucei* genome, alongside the duplex DNA hairpin.

3.3.1.1 Ligand screening

The results of the ligand/G4 FRET screen are shown in Tables 3.1-3.5 and representative curves for F21T-K^+ G4 and F10T duplex DNA are shown in Figure 3.4 (additional data are shown in the Appendix A3, Figures A3.1-A3.3). These data demonstrate the superiority of the *N*-methylpyridinium ligands *E*-**56**, *Z*-**56** and *E*-**57** compared to the first-generation amine derivatives *E*-**41** and *Z*-**41** introduced in Chapter 2. Indeed, whilst only modest stabilisation (ΔT_m = 12 °C) of F21T-K^+ G4 was achieved with ligand *E*-**41** even at high ligand concentration (10 μM), the *N*-methylpyridinium analogue of this compound (*E*-**56**) induces significantly higher thermal stabilisation (ΔT_m = 21 °C, Table 3.1) at lower ligand concentration (1 μM).

Considering the FRET data as a whole, ligand *E*-**56** emerges as the lead compound of this series, inducing the highest thermal stability in all G4 sequences. As in the case of the amine-derived analogue *E*-**41**, the most effective stabilisation was observed for the F21T-K^+ G4 (ΔT_m = 21 °C at 1 μM, Table 3.1). More modest stabilisation values were observed for the F21T-Na^+ G4 (ΔT_m= 12 °C, Table 3.2), FmycT G4 (ΔT_m = 16 °C, Table 3.3) and Febr1T G4 (ΔT_m = 13 °C, Table 3.4). These results suggest that the preference of the stiff-stilbene core for the hybrid G4 topology observed in the first-generation compounds (Chapter 2) is retained to some degree in the new compound series.

The effect of ligand structure on activity was next examined. Whilst *Z*-**56** demonstrates some stabilisation of the G4 sequences, the affinity appears to be significantly lower than for the *E*-**56** counterpart. In general, the thermal stabilisation values induced in all G4 by ligand *Z*-**56** are at least two-fold lower than those observed for *E*-**56** (e.g. at 1 μM, ΔT_m F21T-K^+ = 21 °C for *E*-**56**, 10 °C for *Z*-**56**, Table 3.1). Therefore, the effect of stilbene configuration observed for the first-generation compounds (where the more planar *E* stilbene configuration rendered the ligand more active) appears to be retained in the second-generation *N*-methylpyridinium compounds.

Table 3.1: F21T-K^+ FRET thermal stabilisation values for ligands *E*-**56**, *Z*-**56** and *E*-**57**

	ΔT_m / °C	
Ligand	2 μM	1 μM
E-**56**	26 ± 0.3	21 ± 0.3
Z-**56**	13 ± 0.6	10 ± 0.5
E-**57**	21 ± 0.6	15 ± 0.8

Table 3.2: F21T-Na^+ FRET thermal stabilisation values for ligands *E*-**56**, *Z*-**56** and *E*-**57**

	ΔT_m / °C	
Ligand	2 μM	1 μM
E-**56**	19 ± 0.7	12 ± 0.6
Z-**56**	7 ± 1.3	5 ± 1.0
E-**57**	11 ± 1.2	6 ± 0.9

Table 3.3: FmycT FRET thermal stabilisation values for ligands *E*-**56**, *Z*-**56** and *E*-**57**

	ΔT_m / °C	
Ligand	2 μM	1 μM
E-**56**	23 ± 1.1	16 ± 0.8
Z-**56**	10 ± 0.9	5 ± 1.4
E-**57**	23 ± 0.3	16 ± 0.6

Table 3.4: Febr1T FRET thermal stabilisation values for ligands *E*-**56**, *Z*-**56** and *E*-**57**

	ΔT_m / °C	
Ligand	2 μM	1 μM
E-**56**	19 ± 0.1	13 ± 0.5
Z-**56**	6 ± 0.4	4 ± 0.3
E-**57**	15 ± 0.2	7 ± 0.1

Table 3.5: F10T FRET thermal stabilisation values for ligands *E*-**56**, *Z*-**56** and *E*-**57**

	ΔT_m / °C	
Ligand	2 μM	1 μM
E-**56**	2 ± 0.2	1 ± 0.2
Z-**56**	1 ± 0.1	0 ± 0.1
E-**57**	1 ± 0.3	0 ± 0.2

Comparing *E*-**56** with *E*-**57**, it is interesting to note that whilst both molecules induce high thermal G4 stabilisation, ligand *E*-**56** appears more effective. Indeed, whilst ligand *E*-**56** induces a thermal stabilisation of 21 °C for F21T-K^+ at 1 μM, the stabilisation of ligand *E*-**57** is 6 °C lower (ΔT_m = 15 °C) under comparable conditions (Table 3.1). This effect appears to be general to the majority of the G4 sequences, with the exception of the Febr1T sequence where comparable ΔT_m values are observed for the two ligands (Table 3.4). Therefore, it suggests that the more compact S-shape geometry of ligand *E*-**56** is beneficial to G4 affinity when compared with the more linear rod-like arrangement of ligand *E*-**57**. It is notable that *E*-**56** is more effective at stabilising the G4 sequences despite the DFT calculations suggesting this molecule is not fully planar, in contrast to ligand *E*-**57**. Superficially, this result could be expected on the basis that the dimensions of the S-shaped ligand are better suited to overlap with G-tetrads than the linear counterpart *E*-**57** (Figure 3.3). However, from the FRET results alone, it cannot be discounted at this stage that ligands *E*-**56** and *E*-**57** interact with G4 through mechanisms other than end-stacking, such as groove binding, and further investigations into the binding mode of the ligands will be later discussed.

3.3.1.2 G4/duplex selectivity

Importantly, in addition to the second-generation ligands displaying markedly more potent G4 binding activity than the amine-decorated ligands discussed in Chapter 2, the selectivity for G4 over duplex DNA appears to be retained. Indeed, the ligands induce virtually no stabilisation in the duplex DNA model (ΔT_m < 3 °C, Table 3.5). In conjunction with the results of Chapter 2, these data provide further evidence that stiff-stilbene core appears a promising chemotype for the development of G4-selective binding molecules.

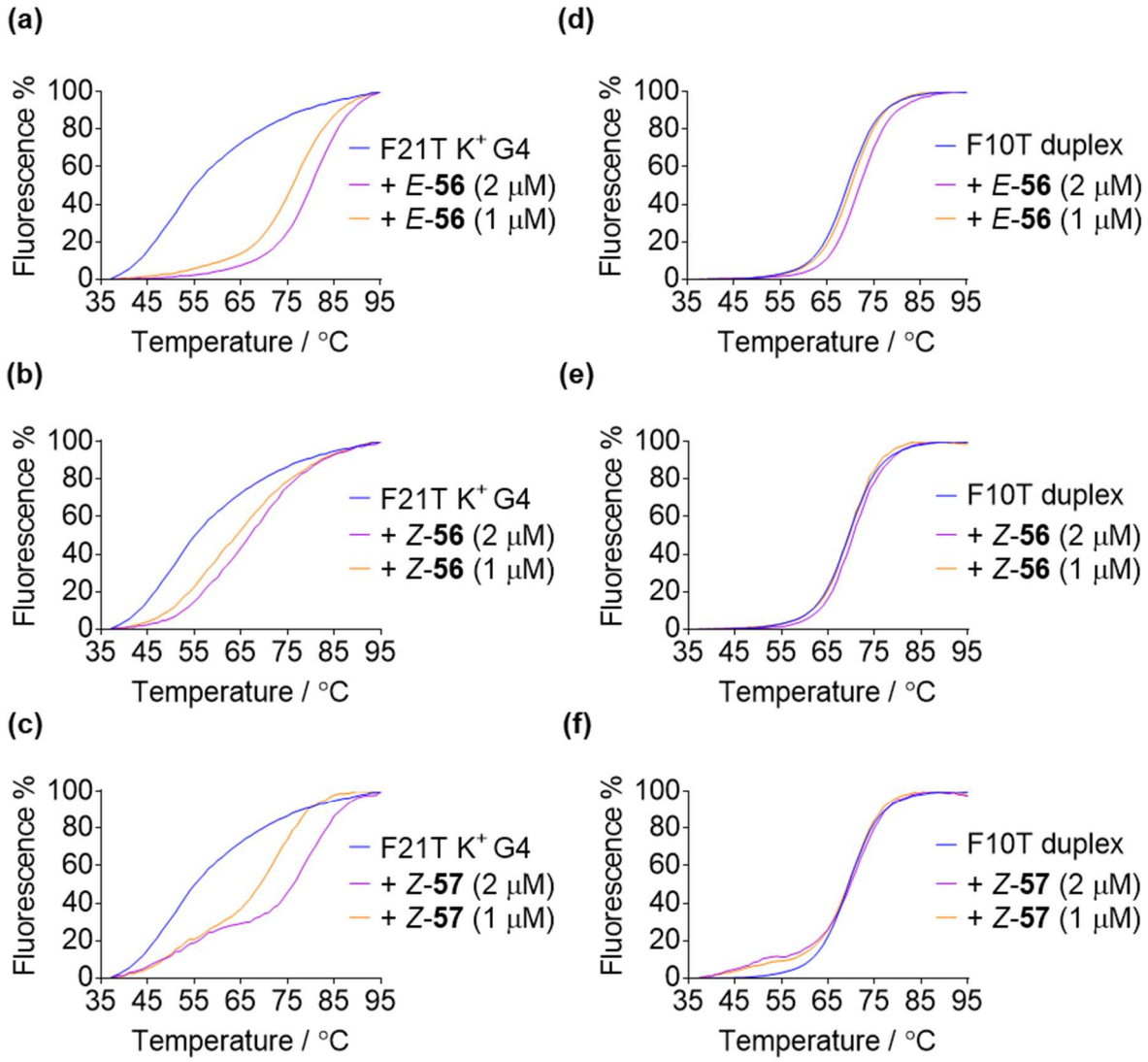


Figure 3.4: Representative FRET curves for (a)-(c) F21T-K⁺ G4 and (d)-(f) F10T duplex in presence of 1 μM and 2 μM ligands (a),(d) *E*-**56**; (b),(e) *Z*-**56**; (c),(f) *E*-**57**.

To further interrogate the G4/duplex selectivity of the pyridinium compounds, competition assays were conducted in which the thermal stabilisation induced in the G4 structures at 1 μM ligand concentration was monitored under conditions of increasing concentration of a non-fluorescent duplex DNA competitor. As mentioned previously (Section 2.3.1), off-target binding is detected by a depression in the induced thermal destabilisation caused by ligand association to the duplex competitor diminishing the concentration of the G4/ligand complex. Results of the competition assays are displayed in Figure 3.5. Pleasingly, the selectivity of the lead compound *E*-**56** for the G4 structures is confirmed (representative melting curves shown for F21T-K⁺ in Figure 3.5a, data represented as black bars in Figures 3.5b-3.5e). At a concentration of 5 μM duplex competitor (25 molar equivalents of duplex to G4 strands), over

70% of the induced thermal stabilisation is retained against F21T-K^+ (Figure 3.5b) and FmycT G4 (Figure 3.5d), demonstrating preferential association with the 4-stranded hybrid and parallel G4 structures. The fact that a measurable depression in the melting temperature is observed indicates however that the discrimination is not complete, unlike for some of the best G4 ligands reported to date, such as PhenDC3 (**10**, Figure 1.14) which resists duplex competitors almost totally up to 10 μM concentration.[130] Nonetheless, the observed results represent a promising selectivity that could be optimised in the development of future generations of stiff-stilbene G4-targeting compounds.

Meanwhile, the discrimination between the F21T-Na^+ antiparallel G4 and duplex appears much less effective (Figure 3.5b). At 5 μM duplex competitor, only ~ 30% stabilisation of the G4 is retained compared to that observed in the absence of competitor. This reflects the weaker affinity for this G4 model indicated by the lower ΔT_m values obtained for F21T-Na^+ (Table 3.2). The selectivity between FebrT G4 and duplex is as also expected based on the initial FRET results (Figure 3.5e). A similar selectivity profile is observed for ligand *E*-**57** (Figures 3.5b-3.5e, dark-grey bars), though the discrimination between G4 and duplex appears to be slightly poorer than for compound *E*-**56**, consistent with this compound being the weaker of the two ligands.

Finally, the *Z*-**56** ligand displays the weakest discrimination between the G4 and duplex DNA structures (Figures 3.5b-3.5e, light-grey bars). Though the selectivity for hybrid G4 appears comparable to *E*-**56** and *E*-**57**, the ligand-induced thermal stabilisation for the all other G4 sequences is clearly depressed even at low concentrations of competitor. This effect is particularly noticeable in the experiments involving FmycT (Figure 3.5d), where a gradual reduction ΔT_m is observed on increasing the concentrations of duplex competitor, with only 19% stabilisation retained under these conditions at 5 μM duplex (compared to the 74% stabilisation retained for the *E*-**56** counterpart under comparable conditions).

The results of the FRET experiments indicate the *N*-methylpyridinium ligands outperform the amine-derived ligands reported in the previous Chapter 2. Fortunately, the desired activity difference between the *E* and *Z* ligand forms (that suggested activity could eventually be controlled by photoisomerisation) was retained and G4/duplex discrimination for the most active compound remained good, demonstrating that increasing the ligand potency did not come at great expense to selectivity. Having identified this new ligand class, subsequent work focused on a more in-depth investigation of the ligand binding modes of the three compounds alongside evaluation of their potential as therapeutic agents using the battery of techniques introduced in the previous chapter.

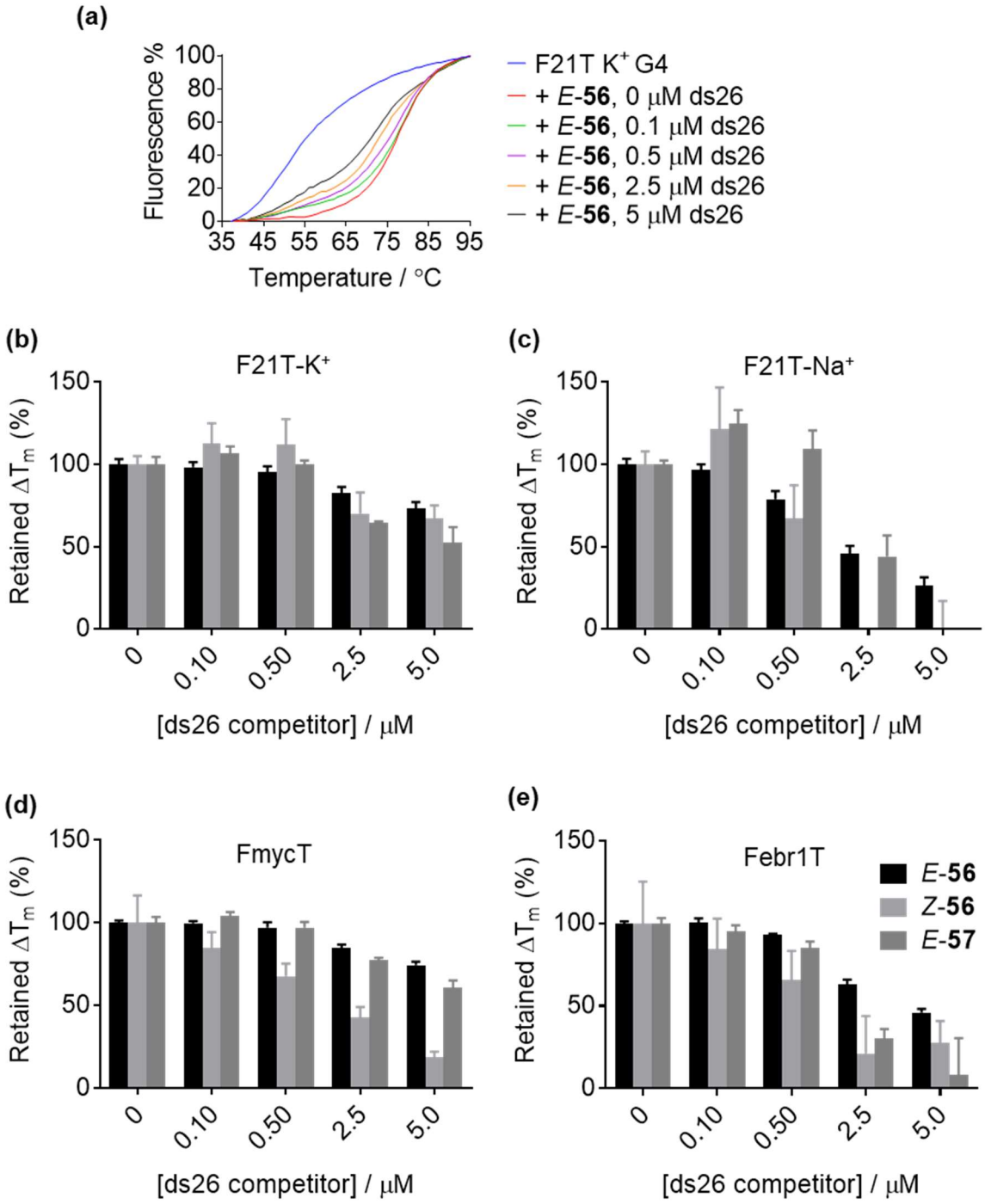


Figure 3.5. FRET competition assay data. (a) Representative melting curves for F21T-K^+ in the presence of *E*-**56** (1 μM) and increasing concentrations of ds26 duplex DNA; (b)-(e) retained thermal stabilisation induced by ligands *E*-**56**, *Z*-**56** and *E*-**57** (1 μM) under increasing concentrations of ds26 duplex DNA for (b) F21T-K^+ G4; (c) F21T-Na^+ G4; (d) FmycT G4; (e) Febr1T G4 sequences.

3.3.2 UV/visible absorbance titrations

Quantification of the binding affinity of ligands *E*-**56**, *Z*-**56** and *E*-**57** to hybrid telo23-K^+ G4 was attempted using UV/visible titration studies in a similar manner as described in Chapter 2 for the amine-derived ligands (Section 2.3.2). The resulting titrations and representative isotherms are shown in Figure 3.6 in which the absorbance spectrum of the ligand was measured upon titration with telo23-K^+ G4. Qualitatively, the changes in the isotherms confirm the strong association of ligands *E*-**56** and *E*-**57** to G4. The isotherms were not well-described by the 1:1 binding model presented in Chapter 2 (Section 2.3.2). Indeed, saturation of binding for ligands *E*-**56** and *E*-**57** is clearly observed before one stoichiometric equivalent of DNA is added (Figures 3.5b and 3.6c) indicating higher binding stoichiometries. Binding of more than one ligand per host quadruplex is more challenging to model owing to co-operative effects, where the binding of the first ligand molecule to the DNA influences the binding of subsequent molecules.[186] Whilst software is available to model such systems, the large number of fitting parameters in the associated equations requires very high-quality, information-rich isotherms in order to obtain reliable fits. Unfortunately, attempts to fit the data that could be obtained to co-operative binding models gave nonsensical results in most cases. The apparent dissociation constants were therefore estimated by fitting the observed isotherms in Figure 3.6 to an independent-and-equivalent sites model, Eqn (3.1) as previously employed by others in the treatment of multi-site G4 ligand binding.[205,206] *N*, the stoichiometry of the complex is the assumed number of ligand binding sites per G4, and was chosen as the lowest value that adequately described the data, with the quality of fit indicated by the R^2 value.

$$\Delta A = \varepsilon_{\Delta D.L} \cdot \frac{\left(N \cdot [D] + [L] + \frac{1}{K_a}\right) - \sqrt{\left(N \cdot [D] + [L] + \frac{1}{K_a}\right)^2 - 4 \cdot N \cdot [D][L]}}{2} \qquad (3.1)$$

Though these assumptions are likely rather crude, the apparent dissociation constants (Figure 3.6) mirror the trends observed in the thermal melting assays with compound *E*-**56** emerging again as the most potent G4 ligand (K_a = 3.8 x 10^6 M^{-1}). The 3:1 binding stoichiometry of this ligand was also supported by circular dichroism studies (see below, Figure 3.7a) as ligand *E*-**56** induces increasing hyperchromicity in the telo23-K^+ bands up to a 3:1 ligand:G4 ratio. These results reinforce the observations of the FRET assay that stilbene configuration is important for G4 recognition, with *Z*-**56** displaying lower apparent affinity (K_a = 1.7 x 10^5 M^{-1}) than the *E*-**56** counterpart, by an order of magnitude.

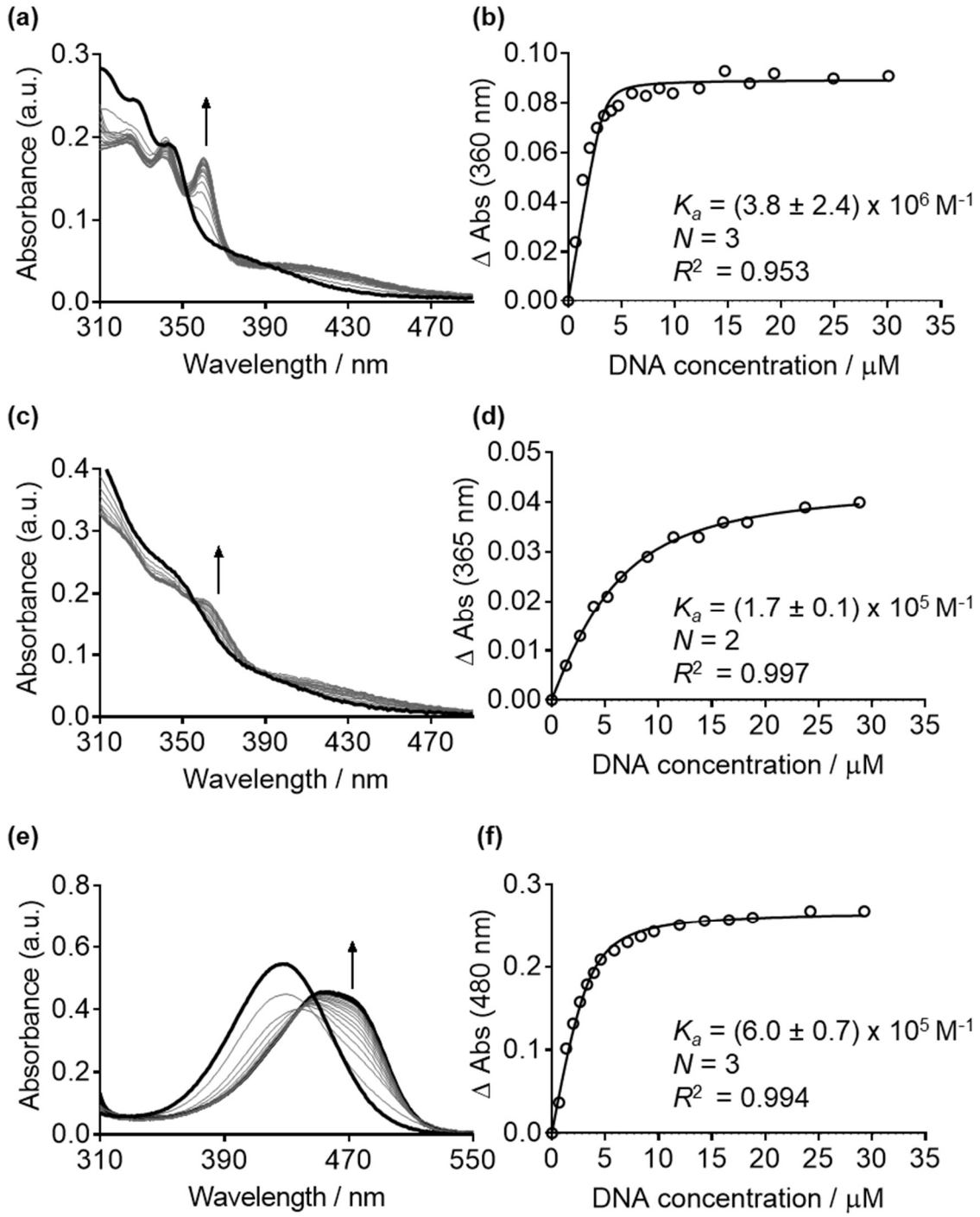


Figure 3.6: G4 binding studies of ligands (a,b) *E*-**56**, (c,d) *Z*-**56** and (e,f) *E*-**57** monitored by UV/visible absorbance titrations. (a,c,e): spectra of 10 μM ligand (black traces) titrated with increasing concentration of telo23-K^+ hybrid G4 (grey traces); (b,d,f) determination of the apparent dissociation constants by fitting the extracted isotherms to Eqn. (3.1) using the stoichiometry indicated.

3.3.3 Circular dichroism titrations

As previously mentioned, circular dichroism (CD) is a useful technique for the study of G4/ligand complex formation, since the effects of the ligand on G4 topology can be detected through changes in the characteristic dichroic absorbance bands. Moreover, induction of CD signals in otherwise CD-transparent systems on association with a chiral macromolecule provides a clear indication of binding. To this end, CD titrations were performed for ligands *E*-**56**, *Z*-**56** and *E*-**57** against different G4 models: the telo23 sequence in both sodium- and potassium-containing buffer (hybrid G4), telo22 in sodium buffer (antiparallel G4) and Pu27 (parallel G4).

This section discusses the differences in CD spectral perturbations induced in the respective G4 signatures by each ligand. In certain cases, these differences are sufficiently striking to confidently conclude that the ligand induces a global change in the DNA conformation. However, there are several instances where more minor changes occur. Though these changes are significant and possible binding modes may be ventured by comparison with data reported for previous systems, the complex origin of CD signals means that it is not possible to draw robust conclusions on the structures of the complexes from CD data alone, since effects are difficult to generalise between different chemotypes. Therefore, in addition to analogy with previous reports, these discussions are also informed by the information gained from NMR spectroscopy and molecular dynamics simulations (Section 3.3.4). Since interpretation of NMR spectra of G4/ligand complexes is itself also complicated by several factors (such as exchange broadening of signals), the combination of both techniques is needed to build evidence for the proposed ligand binding modes.

Examining first ligand *E*-**56**, strikingly different behaviour can be observed between the different G4 topologies examined (Figure 3.7). No major conformational change appears to be induced in the hybrid G4 conformation of telo23 under potassium conditions (Figure 3.7a), though binding of the ligand is clearly evidenced by the ligand-induced hyperchromicity in the positive band at 290 nm consistent with stabilisation of the native hybrid G4. Plotting the ellipticity against the equivalents of ligand *E*-**56** added indicates a 3:1 binding stoichiometry (Figure 3.8). Small induced circular dichroism (ICD) signals are visible in the ligand region ($\lambda > 320$ nm). The origin of ICD signals is complex and difficult to generalise, but small ICD bands have previously been attributed to arise from interaction with the 3-dimensional G4 grooves rather than by end stacking with the planar G-tetrads.[106] Subsequent NMR experiments provided additional evidence that association of ligand *E*-**56** occurs through a groove-binding mode of interaction. Ligand *E*-**56** induces little perturbation to the spectral signature of the parallel Pu27 G4 and no induced CD bands can be detected (Figure 3.7d). This could indicate an alternative mode of binding to the parallel structure compared with the hybrid type.

However, in the absence of further structural information regarding binding to the parallel sequence (not obtained in the present study) this inference is rather speculative.

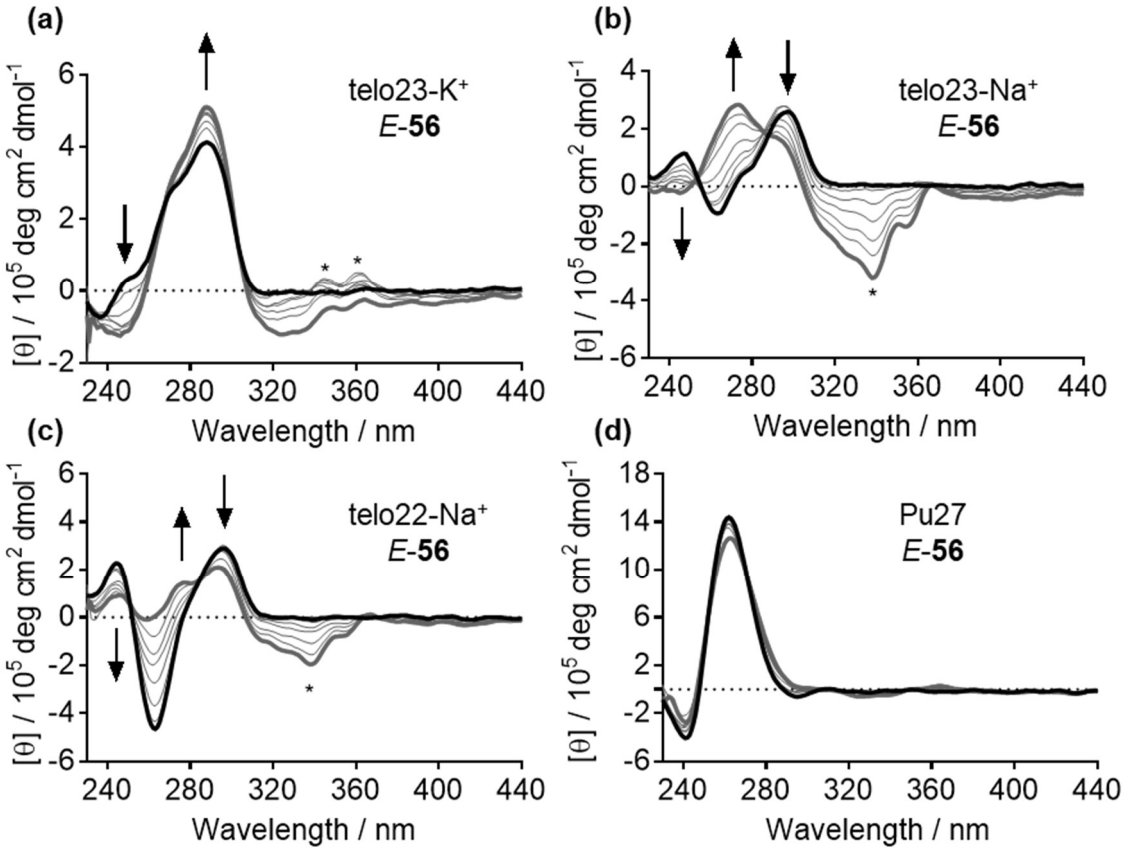


Figure 3.7. CD titrations of studies of ligand *E*-**56** against (a) telo23-K^+ (hybrid G4); (b) telo23-Na^+ (hybrid G4); (c) telo22-Na^+ (antiparallel G4); (d) Pu27 (parallel G4). Black trace: DNA CD spectra in absence of ligand; dark grey trace: CD spectra in the presence of 7 equiv. *E*-**56**. Intermediate titration points are shown as pale grey traces. [θ] = molar ellipticity. DNA concentration = 4.2 μM.

Of perhaps greater interest is the effect of ligand *E*-**56** on the G4 sequences examined in sodium buffer (Figure 3.7b). Though the CD spectrum of telo23 in sodium-rich conditions is different from the same sequence in potassium-containing buffer (Figure 3.7b), it was recently demonstrated by detailed structural investigations that the same hybrid topology of telo23 is observed under Na^+ and K^+-rich conditions.[207] However, despite the similarity in G4 secondary structure in the absence of ligand, the effects of ligand *E*-**56** on the DNA topology are markedly different.

Upon titration with ligand *E*-**56**, all of the characteristic bands associated with the native G4 fold of telo23-Na^+ disappear and the emergence of new spectral features can clearly be seen, specifically a positive band at 270 nm and a large induced CD signal in the ligand region (> 320 nm). These changes indicate the ligand disrupts the G4 fold and induces rearrangement to an alternative structure. The CD spectra of the telo23-Na^+/*E*-**56** complex does contain any features corresponding to archetypal G4 signatures and, on the basis of the CD data, all that can be concluded is that significant structural perturbation is caused by the ligand. However, subsequent NMR experiments and computational studies provide further evidence that the ligand induces denaturation of the G4 folding under these conditions (Section 3.3.4).

Similar spectral changes are induced in the telo22 sequence in the same buffer conditions, which forms instead an antiparallel G4 topology (Figure 3.7c). However, the effect is much less pronounced than in the case of telo23, and at the final titration point the CD spectrum reveals a mixture of antiparallel G4 and the remodelled species. Specifically, the maximum at 290 nm can still be observed in addition to ligand-induced positive band at 273 nm; in the titration of telo23-Na^+, the positive band at 290 nm is barely observed at the end of the experiment.

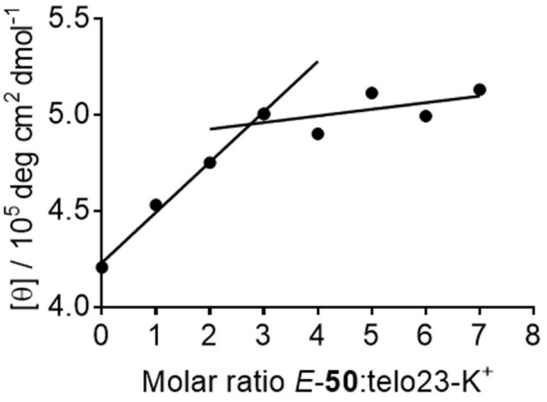


Figure 3.8: Molar ellipticity at 290 nm of telo23-K^+ hybrid G4 upon titration with ligand *E*-**56**, indicating 3:1 binding stoichiometry.

To provide further evidence that the spectra changes induced by ligand *E*-**56** were indeed the result of global conformational rearrangement, a kinetic experiment was performed (Figure 3.9). The emergence of the new positive feature at 270 nm was monitored following the mixing of the telo23-Na^+ G4 and 10 equiv. ligand *E*-**56**. This revealed the spectral changes to take place in the order of minutes, suggesting a global structural rearrangement. A characteristic unfolding time (τ) of 245 seconds was extracted from fitting of the data to the first-order rate

Eqn. (3.2). This timescale is consistent with ligand-induced G4 topological changes reported by others. For example, similar rearrangement times were reported by Chaires and co-workers for the TMPyP4-mediated conformational switch of telomeric DNA.[208]

$$\theta = a - be^{-t/\tau} \tag{3.2}$$

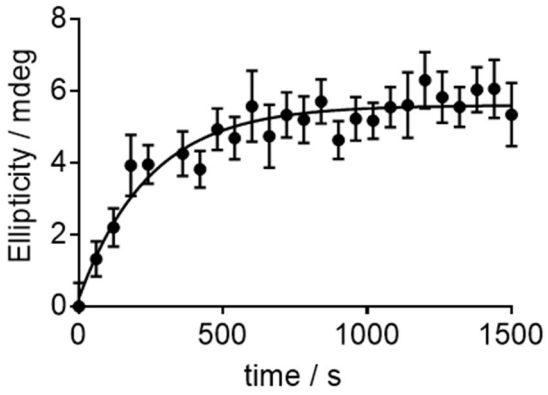


Figure 3.9: CD kinetic study of telo23-Na^+ refolding after addition of 10 equiv. ligand *E*-**56**, monitoring the emergence of the positive feature at 270 nm with time.

Next, the effect of the *Z*-**56** ligand on the various G4 structures was examined under analogous conditions (Figure 3.10). The ligand interacts very differently with the G4 topologies compared to *E*-**56**. Considering first the telo23-K^+ hybrid G4, whereas titration of the quadruplex with *E*-**56** preserves and increases the intensity of the characteristic G4 bands, addition of ligand *Z*-**56** causes attenuation of the shoulder band at 260 nm characteristic of the native equilibrium (Figure 3.10a) and, unlike for ligand *E*-**56**, no ICD signals are observed in the ligand region. Subsequent NMR experiments (Section 3.3.4) provided further evidence that the CD spectral changes result from ligand-induced perturbation of the conformational equilibrium of G4. Only very minor changes are seen in the spectrum of the parallel G4 structure (Figure 3.10d), though a small induced CD in the ligand region (330 nm) indicates binding between the two structures.

Examining the two G4s formed in sodium-rich conditions, interesting differences are again observed, both between telo22 and telo23 G4 in this buffer and in the behaviour of ligand *Z*-**56** compared to ligand *E*-**56**. Virtually no perturbations to the antiparallel G4 signature are caused by the addition of ligand *Z*-**56** (Figure 3.10c), indicating persistence of the antiparallel

topology. In surprising contrast, the telo23-Na^+ hybrid G4 signature is significantly perturbed by the addition of ligand *E*-**56** and at the end of the titration the spectrum resembles exactly that of the telo22 antiparallel G-quadruplex, suggesting that *Z*-**56** may induce this structure from the hybrid type found in the absence of ligand. Most strikingly, the disruption of the quadruplex structure induced in both telo22-Na^+ and telo23-Na^+ by *E*-**56** is not observed for the *Z*-**56** counterpart; in both cases, a characteristic G4 spectral signature is preserved throughout the titration.

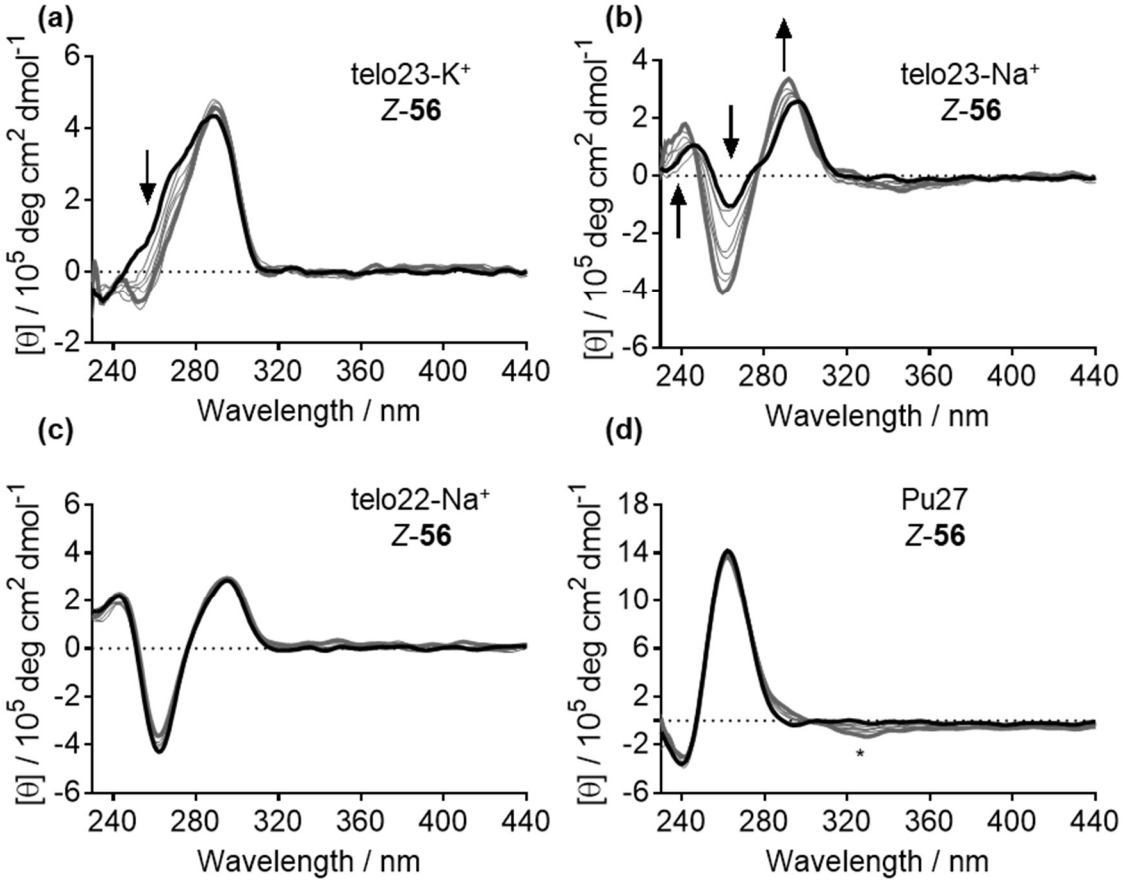


Figure 3.10: CD titrations of studies of ligand *Z*-**56** against (a) telo23-K^+ (hybrid G4); (b) telo23-Na^+ (hybrid G4); (c) telo22-Na^+ (antiparallel G4); (d) Pu27 (parallel G4). Black trace: DNA CD spectra in absence of ligand; dark grey trace: CD spectra in the presence of 7 equiv. *Z*-**56**. Intermediate titration points are shown as pale grey traces. [θ] = molar ellipticity. DNA concentration = 4.2 μM.

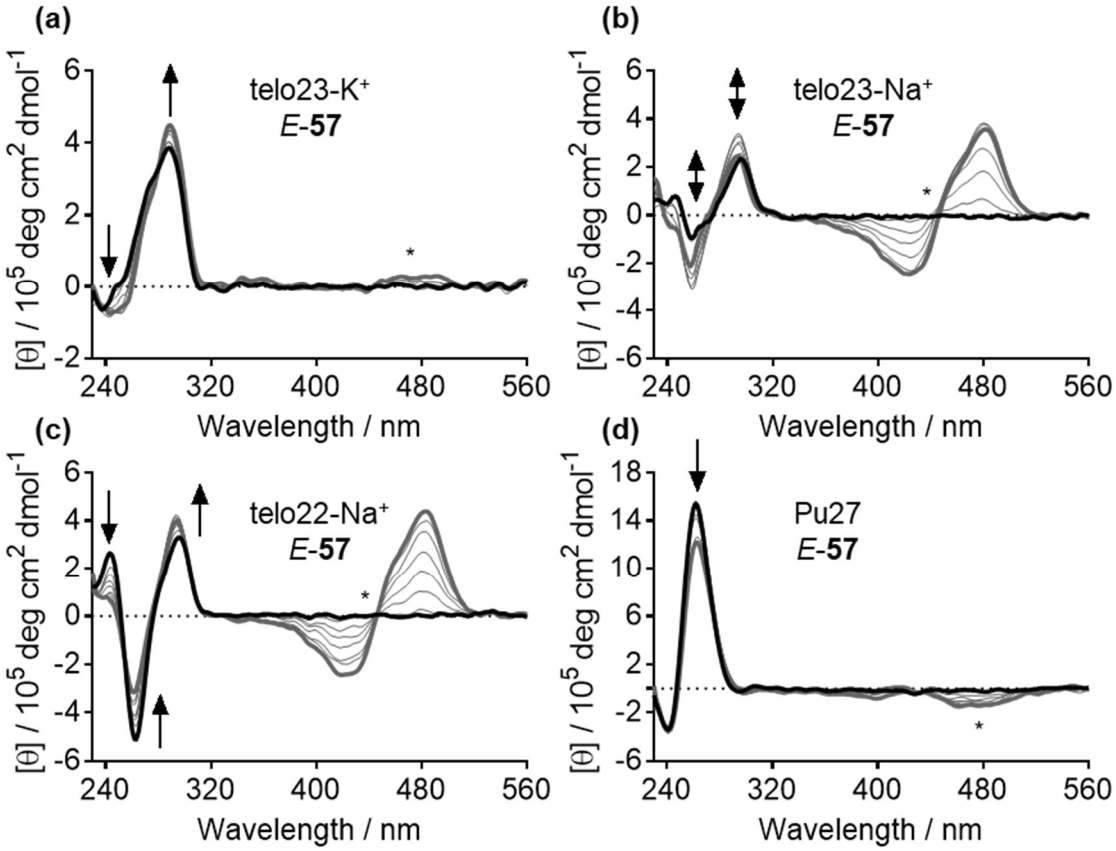


Figure 3.11: CD titrations of studies of ligand *E*-**57** against (a) telo23-K^+ (hybrid G4); (b) telo23-Na^+ (hybrid G4); (c) telo22-Na^+ (antiparallel G4); (d) Pu27 (parallel G4). Black trace: DNA CD spectra in absence of ligand; dark grey trace: CD spectra in the presence of 7 equiv. *E*-**57**. Intermediate titration points are shown as pale grey traces. [θ] = molar ellipticity. DNA concentration = 4.2 μM.

Finally, the activity of ligand *E*-**57** was investigated (Figure 3.11). The binding mode to the hybrid telo23-K^+ G4 appears similar to that for ligand *E*-**56** (Figure 3.11a), with similar spectral changes (namely hyperchromicity at 290 nm and ICD signals in the ligand region) observed during the titration. Interaction of the ligand with the parallel Pu27 sequence is also evidenced by perturbation of the positive band and negative ICD signal in the ligand region (Figure 3.11d). Interestingly, a large bisignate ICD is observed in the ligand in both telo23-Na^+ (Figure 3.11b) and telo22-Na^+ (Figure 3.11c). Conveniently, this blue-shifted ligand absorption band does not overlap with the absorption spectrum of the DNA, so perturbations in the G4 spectral signature can be detected more easily than in the case of *E*-**56**. In the case of telo22-Na^+,

hypochromicity is observed in the G4 bands at 240 nm and 260 nm, indicative of some disruption of the fully folded structure, but the effect is not as pronounced as for ligand *E*-**56**. The behaviour of the hybrid telo23-Na^+ sequence is more complex. Initially, the ligand induces hyperchromicity in both the negative band a 260 nm and positive bands at 290 nm, but at higher ligand concentrations the intensity of these bands becomes depleted, suggesting *E*-**57** is also capable of disrupting the G-tetrads of telo23-Na^+ at higher concentrations.

3.3.4 NMR and molecular dynamics studies

Though the CD experiments are useful to examine qualitative differences in the interaction of the ligands with the different G4 structures, more direct structural information was needed to build more additional evidence of the proposed binding modes and ligand effects on G4 structure. NMR experiments were therefore conducted on the hybrid (telo23-K^+) and antiparallel (telo22-Na^+) for the three ligands to examine the structural effects on these different G4 topologies in more detail.

3.3.4.1 Studies with telo23-K^+ hybrid G4

The binding of the ligands to hybrid G4 in potassium-containing buffer was first investigated. Whilst CD indicated the folded G4 structure was preserved upon the addition of all three ligands *E*-**56**, *Z*-**56** and *E*-**57**, different binding modes were observed between for the *E*-**56** and *Z*-**56** ligands (Section 3.3.3). Furthermore, this is the more biologically-relevant G4 owing to the high intracellular concentration of potassium[209] and the ligands induced the highest stabilisation values in this sequence in the thermal melting assay (Section 3.3.1). As discussed in Chapter 2, the telo23 sequence adopts a mixture of a major and minor folded species under the NMR conditions, meaning only the imino resonances could be straightforwardly assigned due the spectral complexity in the aromatic region. Nonetheless, the behaviour of the imino signals during the titration provides some information about the nature of the G4/ligand interactions to complement the CD data.

Examining first ligand *E*-**56** (Figure 3.12), severe line broadening of the imino resonances can be observed during the titration. All imino signals broaden to a similar degree and remain distinguishable, suggesting that interactions with specific G-tetrad residues do not dominate in the association of ligand *E*-**56** with G4**,** providing additional evidence for the groove-binding mode inferred from the CD titrations. Furthermore, end-stacking with G4 bases would likely result in chemical shift perturbations of specific resonances due to ring-current effects,[106] which are not observed.

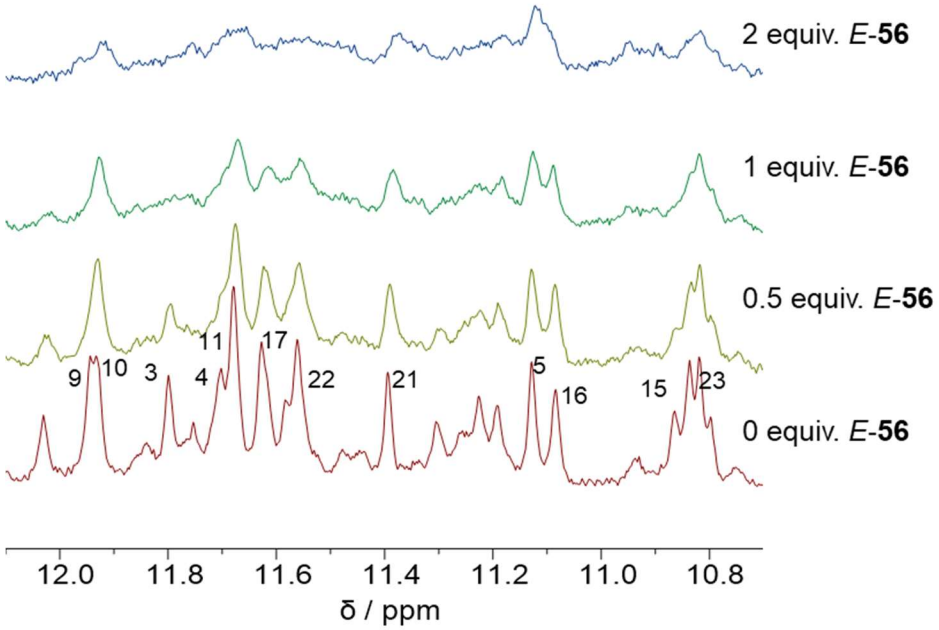


Figure 3.12: NMR titration of telo23-K^+ hybrid G4 with ligand *E*-**56**. DNA concentration = 185 μM.

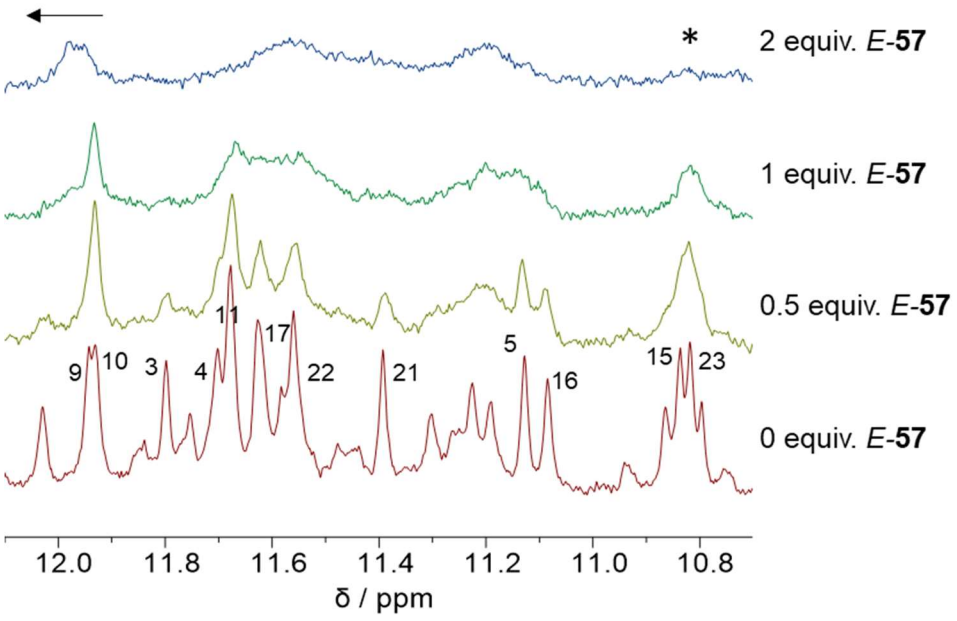


Figure 3.13. NMR titration of telo23-K^+ hybrid G4 with ligand *E*-**57**. The shift in G9/G10 resonances is indicated with an arrow and the disappearance of G15/G23 resonances is indicated *. DNA concentration = 185 μM

Ligand *E*-**57** (Figure 3.13) also induces spectral line broadening at 1:1 stoichiometry though unlike for ligand *E*-**56** the G9/G10 signals shift significantly at 2:1 stoichiometry, whilst signals associated with the lower G-tetrad (G15/G23) disappear entirely, suggesting specific interactions with these parts of the G4, possibly stacking or intercalation, are also important in the binding of ligand *E*-**57**.

For ligand *Z*-**56** (Figure 3.14), all imino resonances are resolved at 2:1 ligand:G4 stoichiometry unlike what was observed for ligands *E*-**56** and *E*-**57**. Notably, whilst the unbound G4 sequence exists as a mixture of major and minor conformations, complexation with ligand *Z*-**56** appears to favour a single conformation, with only 12 distinct imino signals visible at 2:1 stoichiometry versus the more complex spectrum for the G4 sequence observed in the absence of the ligand. This structural perturbation may explain the disappearance of the shoulder band in the CD spectrum of telo23 upon titration with ligand *Z*-**56** (Figure 3.10a).

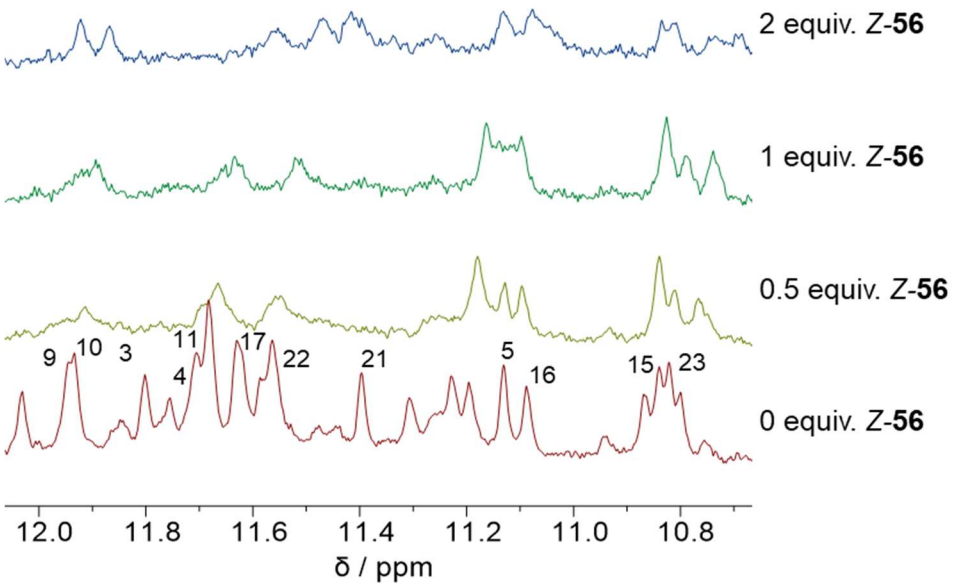


Figure 3.14. NMR titration of telo23-K^+ hybrid G4 (imino region) with ligand *Z*-**56**. DNA concentration = 185 μM.

In conjunction with the data obtained in the CD titrations, it can be inferred that ligand *E*-**56** interacts with hybrid telo23-K^+ G4 primarily through groove binding without association with specific G-tetrad bases, whilst stacking interactions with G-tetrad regions are also important in the binding of ligands *E*-**57** and *Z*-**56**.

It should be stated that though the NMR experiments and CD experiments together provide evidence of certain binding modes, unambiguous determination is not possible based on the current data, and more involved future structural work will be necessary to prove the exact mechanisms of interaction.

*3.3.4.2 Unfolding of G4 in Na^+ conditions by E-***56**

The CD studies (Section 3.3.3) indicated the whilst the hybrid G4 form was stable to the binding of ligand *E*-**56** in potassium-rich conditions (Figure 3.7a), the same ligand induces dramatic changes in G4 folding conformation in sodium-containing buffer (Figure 3.7b). Since the preservation of the G4 fold of telo23-K^+ upon titration with ligand *E*-**56** was also demonstrated by NMR (despite the line broadening, the presence of signals consistent with Hoogsteen-bonded G-tetrads remains clearly visible) the effect of ligand *E*-**56** was next examined against the same sequence in sodium-containing buffer (Figure 3.15a).[207] Upon titration with ligand *E*-**56**, line broadening is initially observed, but at 2:1 stoichiometry, the imino resonances almost completely disappear, indicating the disruption of the Hoogsteen network of the G-tetrads by the addition of ligand. It was concluded that effects observed in the CD studies result from the *E*-**56** induced denaturation of telo23-Na^+ structure. That the aromatic resonances persist, albeit broaden considerably, indicates the loss of imino resonances is not simply a result of precipitation.

The same experiment was attempted on telo22-Na^+ (antiparallel G4) but the evidence for G4 unfolding was not as conclusive as for the telo23-Na^+ hybrid model (Figure 3.15b). Nonetheless, this behaviour correlates with the CD data which indicate that the antiparallel telo22-Na^+ G4 is more resistant to unfolding by ligand *E*-**56** than the hybrid telo23-Na^+ structure (compare Figure 3.7b and 3.7c).

To provide further insight into the unfolding of the G4 structure by ligand *E*-**56**, computational studies were conducted by Dr Susanta Haldar and Adrian Mulholland (Figures 3.16 and 3.17).

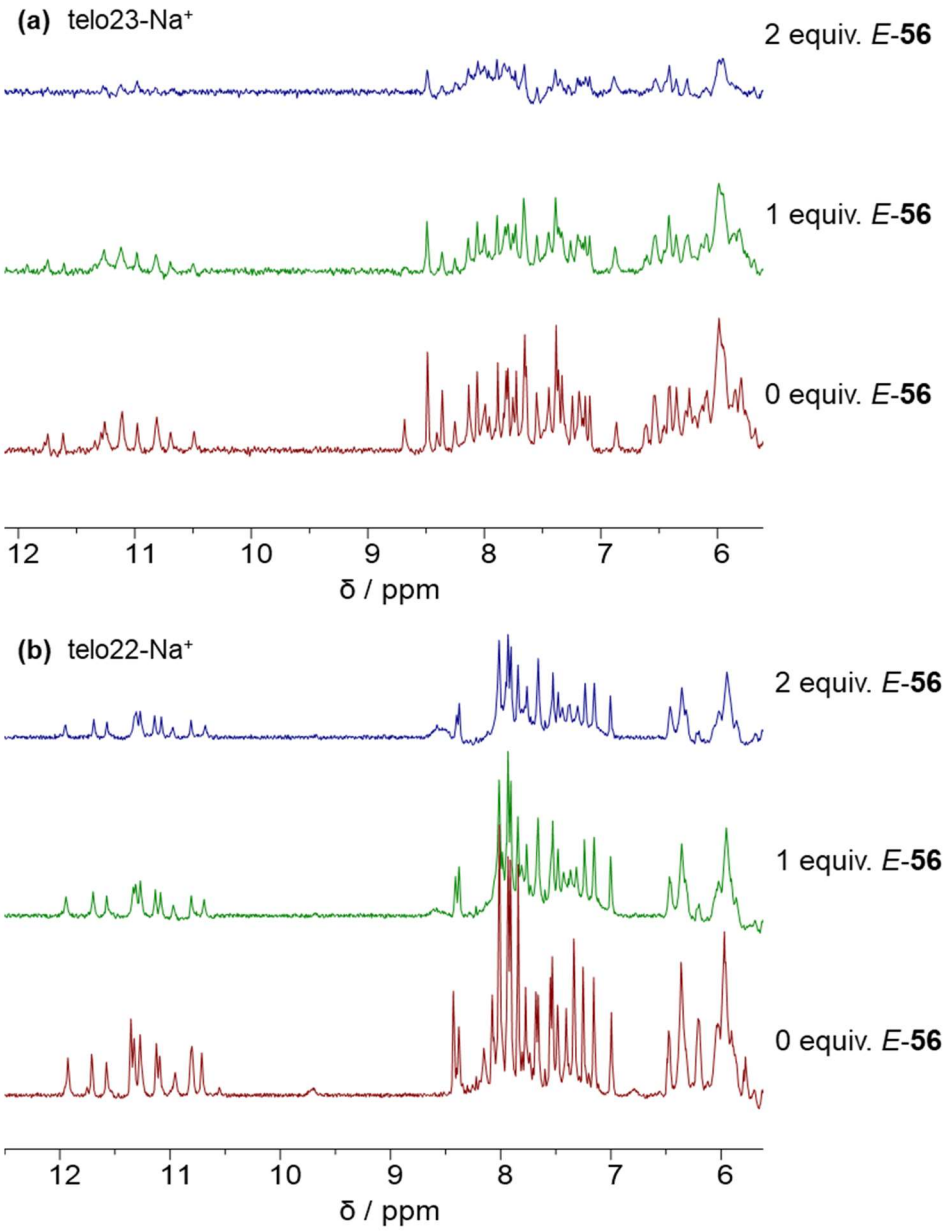


Figure 3.15. NMR titrations of (a) telo23-Na^+ hybrid G4 and (b) telo22-Na^+ antiparallel G4 with ligand *E*-**56**. DNA concentration = 185 μM.

The ligand was first docked to the G4 structure to predict the likely binding poses, and the lowest-energy pose was then subjected to 1 μs molecular dynamics simulations. In these simulations the ligand samples a range of binding poses, which energetically favourable orientations being sampled more frequently, thus indicating the preferred binding orientations (Figure 3.16). In the case of ligand *E*-**56** the MD simulations indicate three preferred ligand

conformations on the top face of the G4 (Figure 3.16a and 3.16c). However, MD simulations do not allow the system access states outside these local minima, such as the alternative binding events that would trigger DNA unfolding. These issues may be addressed by employing well-tempered metadynamics, an advanced modelling technique that biases the system towards exploring rarely sampled configurations (Figure 3.17). This provides a more detailed picture of ligand binding and unbinding events, and allowed longer-term effects, such as conformational remodelling, to be observed.

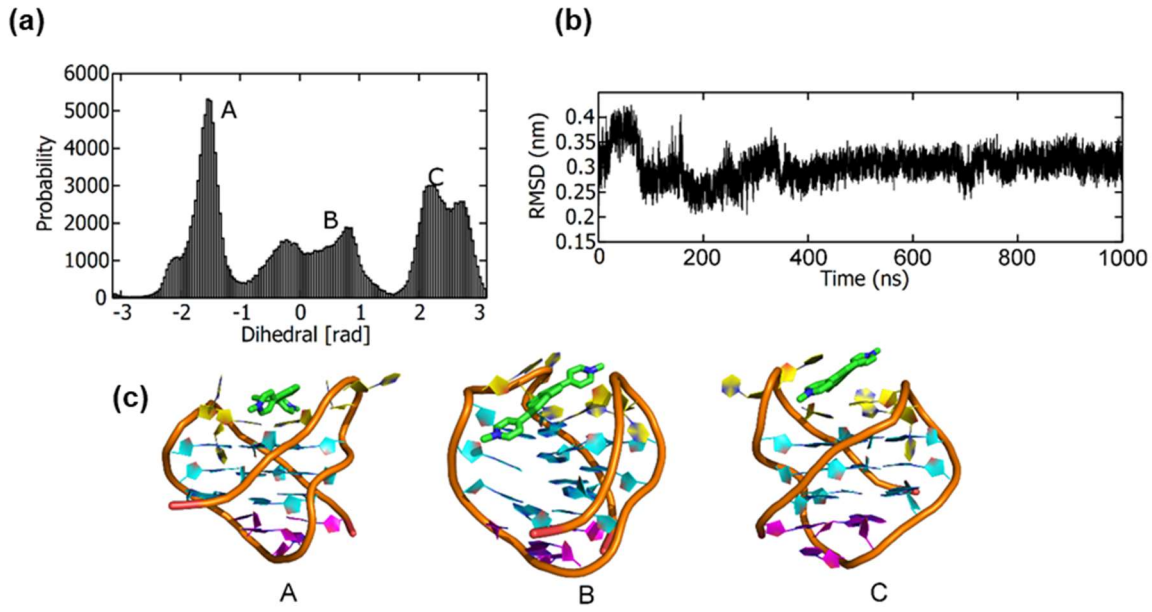


Figure 3.16. Molecular dynamics simulations of telo22-Na^+ and *E*-**56**. (a) probability histogram of ligand orientations sampled, characterised by the dihedral angle between ligand and G4; (b) stability of the G4 backbone co-ordinates over the time course of the simulation (c) binding poses corresponding to maxima A, B and C in Figure (a). *Figures prepared by Dr Susanta Haldar.*

The ligand binding and unbinding events observed by the metadynamics simulations are visible in Figure 3.17a, which plots the separation distance between the centre of mass of the central G-tetrad that of the ligand against the timescale of the simulation. The corresponding molecular visualisation is shown below (Figure 3.17b). From the starting docking pose from the molecular dynamics simulations (Pose A), the ligand unbinds from the G4 within the first ~15 ns and samples the solvated state. To dissociate from the G4, the ligand firsts slides on top of the DNA and interacts with the capping residues (Pose B) which then interacts open towards the solvent (pose C). Finally, the ligand dissociates from these bases and enters the

solvated state (pose D). The process of intercalation and unfolding of the G4 can be followed during the remainder of the simulation.

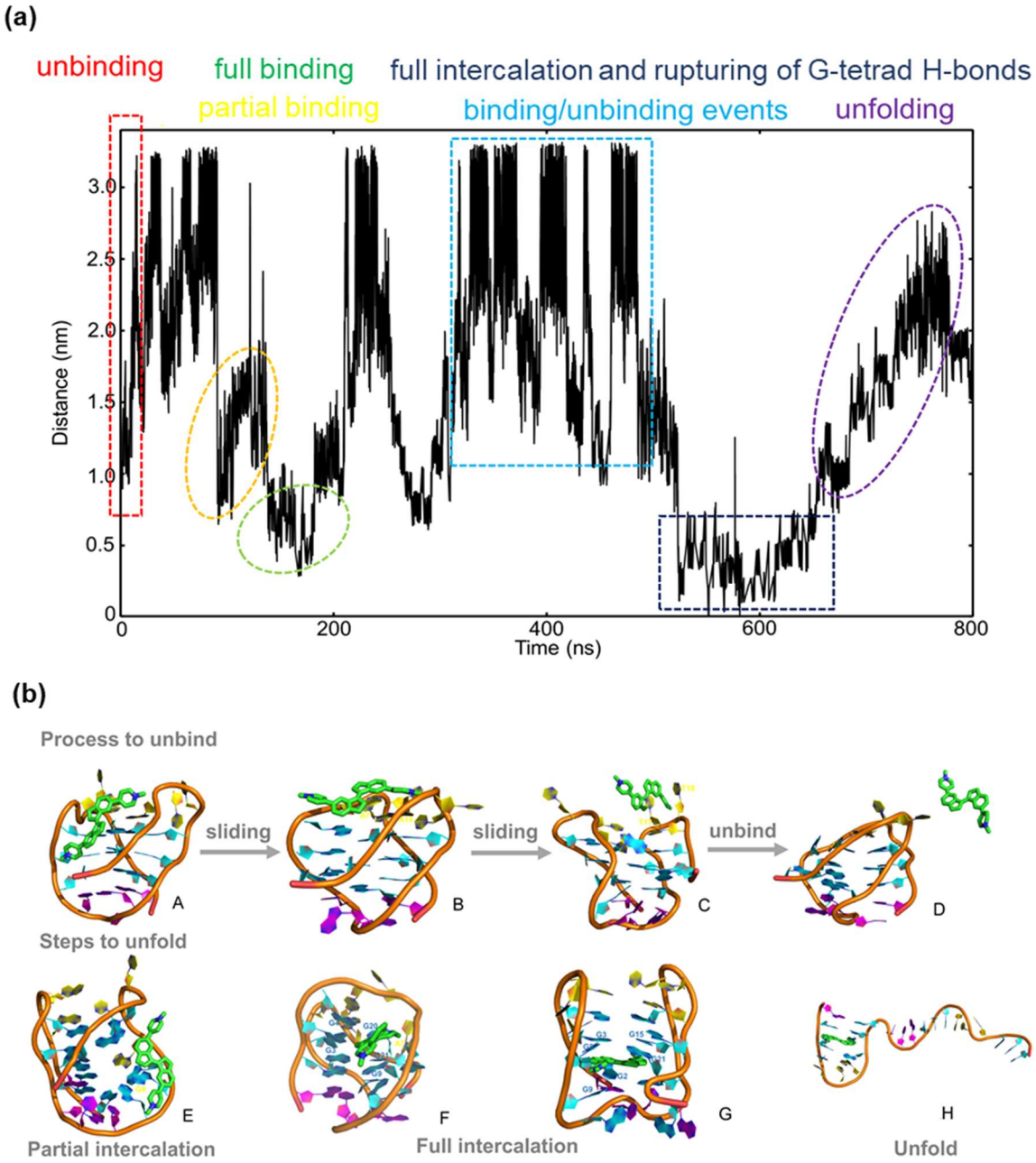


Figure 3.17: Well-tempered metadynamics studies examining the unfolding mechanism of telo22-Na^+ G4 by ligand *E*-**56**. (a) binding/unbinding events characterised by distance between ligand and G4 centres of mass; (b) representative structures corresponding to different stages of the binding/unfolding process. *Figure prepared by Dr Susanta Haldar.*

After 90 ns, the ligand re-enters the groove and then intercalates the G4 structure, partially rupturing the G-tetrads (Pose E). In the later stage of the simulation (150 ns) full intercalation takes place (Pose F). The ligand is then seen to sample bound and dissociated states. At ~525 ns, the ligand again enters the binding pocket and intercalates with G21, G8, G15, G2 and G14 (Pose G). Finally, at ~580 ns, the telo22 G4 begins to unfold within the next 100 ns (pose H). These results provide a plausible mechanism of how the binding mode of ligand *E*-**56** results in the structural perturbations observed in the experimental studies.

*3.3.4.3 Stability of telo22-Na^+ G4 to Z-***56**

Whilst the metadynamics studies supported the unfolding of G4 prompted by ligand *E*-**56** as observed in the CD and NMR experiments, it was also necessary to characterise the binding of the *Z*-**56** ligand to telomeric G4 using the same methods to gain confidence that the simulations accurately represent the behaviour of the G4/ligand systems. Ligand *Z*-**56** did not appear to induce unfolding of either telo23-Na^+ or telo22-Na^+ in the CD studies, and hence should not be seen to induce G4-unfolding in the simulations if the model offers a reliable description of the binding process.

The theoretical calculations were performed by Dr Haldar and Dr Mulholland using the same procedures employed for the *E*-**56** system (Figures 3.18 and 3.19). The molecular dynamics simulations from the initial docking poses indicated a stable binding pose of the ligand at the top of the G4, associating with the A7, T18 and A19 bases on the lateral loops (Figures 3.18a and 3.18c). The overall RMSD of the DNA backbone is found to be constant at ~0.3 nm (Figure 3.18b) indicating that G4 is stable during the MD simulation. Next, the stable binding pose was subjected to well-tempered metadynamics simulations in order to more thoroughly understand the binding process to the G4 and confirm the stability of the folded structure to ligand binding (Figure 3.19). In accordance with the experimental results, the folded G4 is found to be stable to the binding of *Z*-**56** and the simulations fully converge after ~800 ns. This convergence allows the free energy surface (FES) of ligand binding to be computed and information regarding the binding mechanisms can be inferred by inspection of the local energy minima. The FES is shown in Figure 3.18, expressed as a function of d (distance) and T (torsion) collective variables, which reflect the ligand orientation and distance from the centre of mass from the G4. Four principle free energy minima are observed, where Basin A is the global minimum.

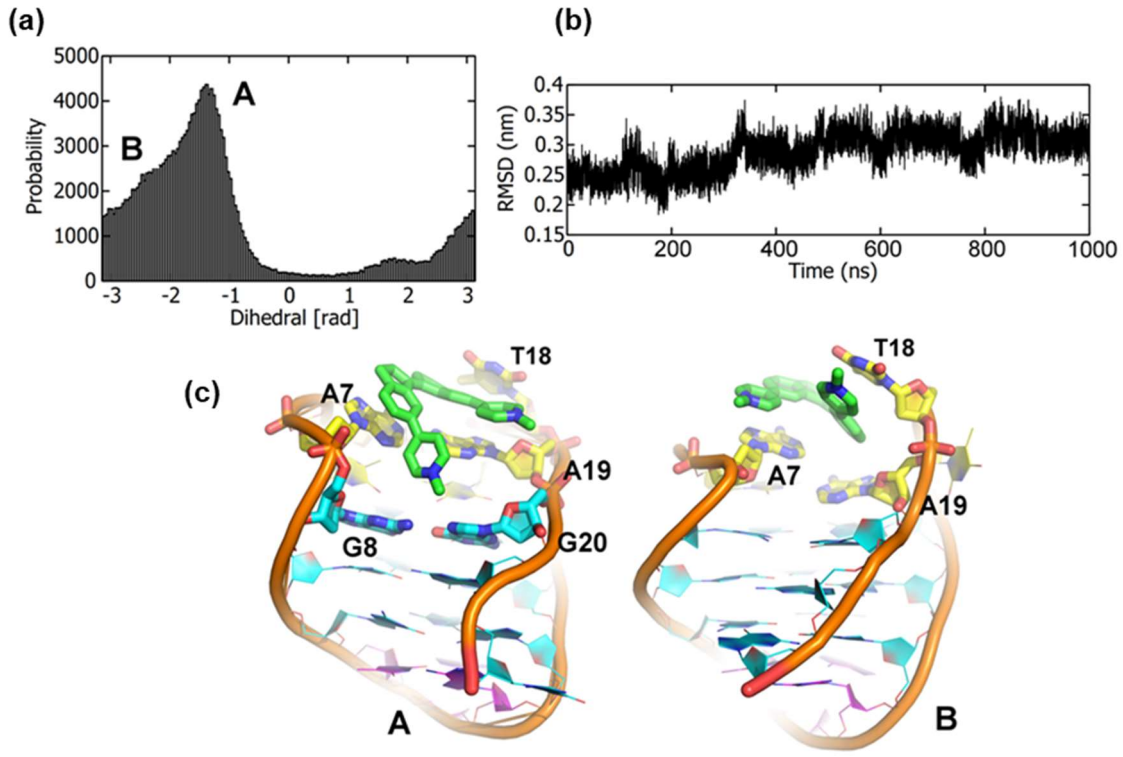


Figure 3.18: Molecular dynamics simulations of telo22-Na^+ and *Z*-**56**. (a) probability histogram of ligand orientations sampled, characterised by the dihedral angle between ligand and G4 axes; (b) stability of the G4 backbone co-ordinates over the time course of the simulation (c) binding poses corresponding to maxima A and B in Figure (a). *Figures prepared by Dr Susanta Haldar.*

In this pose, the ligand interacts with the top face of the G4, by sandwiching in between T6, A7, A19 and T18 bases. In Basin D, one of the pyridinium rings stacks with the A7 base, but both indane rings partially interact with the G8 and G20 bases which belong to the top G-tetrad. In order to sample this binding pose from the global minimum (Basin A) the ligand slides from T18 to interact with the A7 base. These results suggest that the binding of the ligand involves the top face of the G4. In Basin B and C, the ligand is partially exposed to the solvent and therefore these local minima likely represent intermediate states of binding and dissociation from the G4. No local minima were found in the groove regions of the G4 in the WT-MetaD studies.

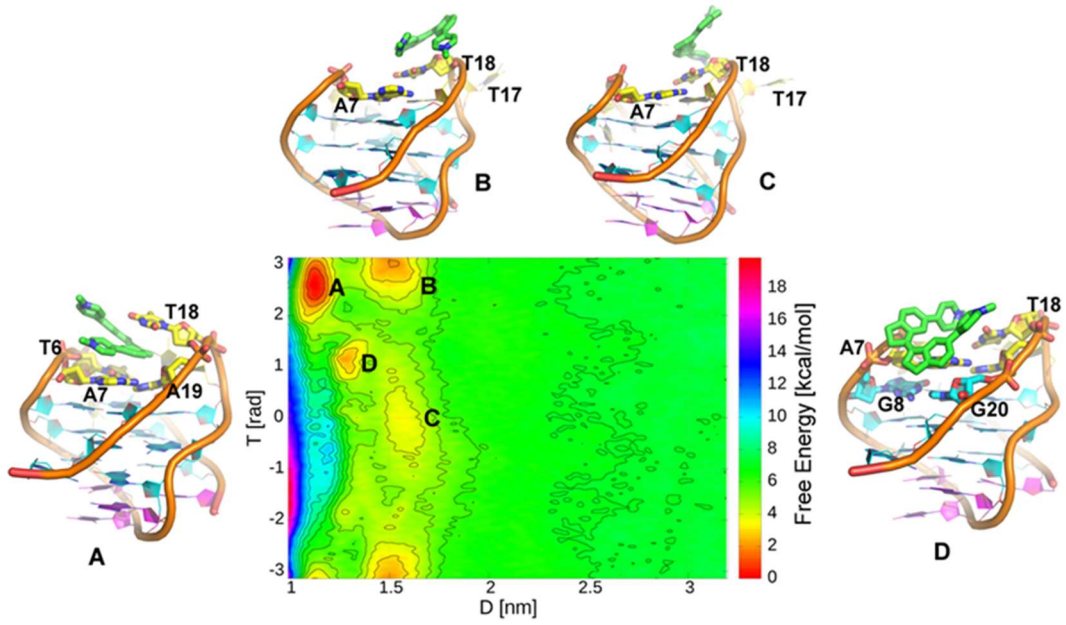


Figure 3.19: Free-energy surface of the telo22-Na^+ / *Z*-**56** binding process computed as a function of ligand distance and torsion angle. Representations of ligand poses in the local minima are shown (Poses A-D).

In order to obtain physical evidence of the binding modes indicated by the calculations, the telo22/*Z*-**56** complex was investigated by NMR spectroscopy (Figure 3.19). Since the binding to this quadruplex is rather weak, a fast-exchange regime is observed where the chemical shift perturbations of the DNA resonances vary gradually throughout the titration (Figure 3.20a), since they appear as the weighted average of the chemical shifts of the free and bound states (see discussion in Section 2.3.4, Figure 2.19). Furthermore, as the secondary structure of the G4 is preserved upon addition of the ligand, the resonances several of the aromatic and ribose protons could be assigned in addition to the imino signals corresponding to the bases of the G-tetrads from 2D NOESY spectra (Figures 3.20b and 3.20c, full spectrum in Appendix A3, Figure A3.4).

Key binding residues can be inferred from the chemical shift perturbations induced by ligand *Z*-**56**. The G20 and G4 imino resonances (Figure 3.20a and Figure 3.20c) and the G16H8 proton (Figure 3.20c), corresponding to the top G-tetrad, are clearly perturbed. Further significant chemical shift perturbations are seen for the A7H1' (anomeric) proton and the A19H8 (aromatic) proton corresponding to residues in the lateral loops (Figures 3.20b).

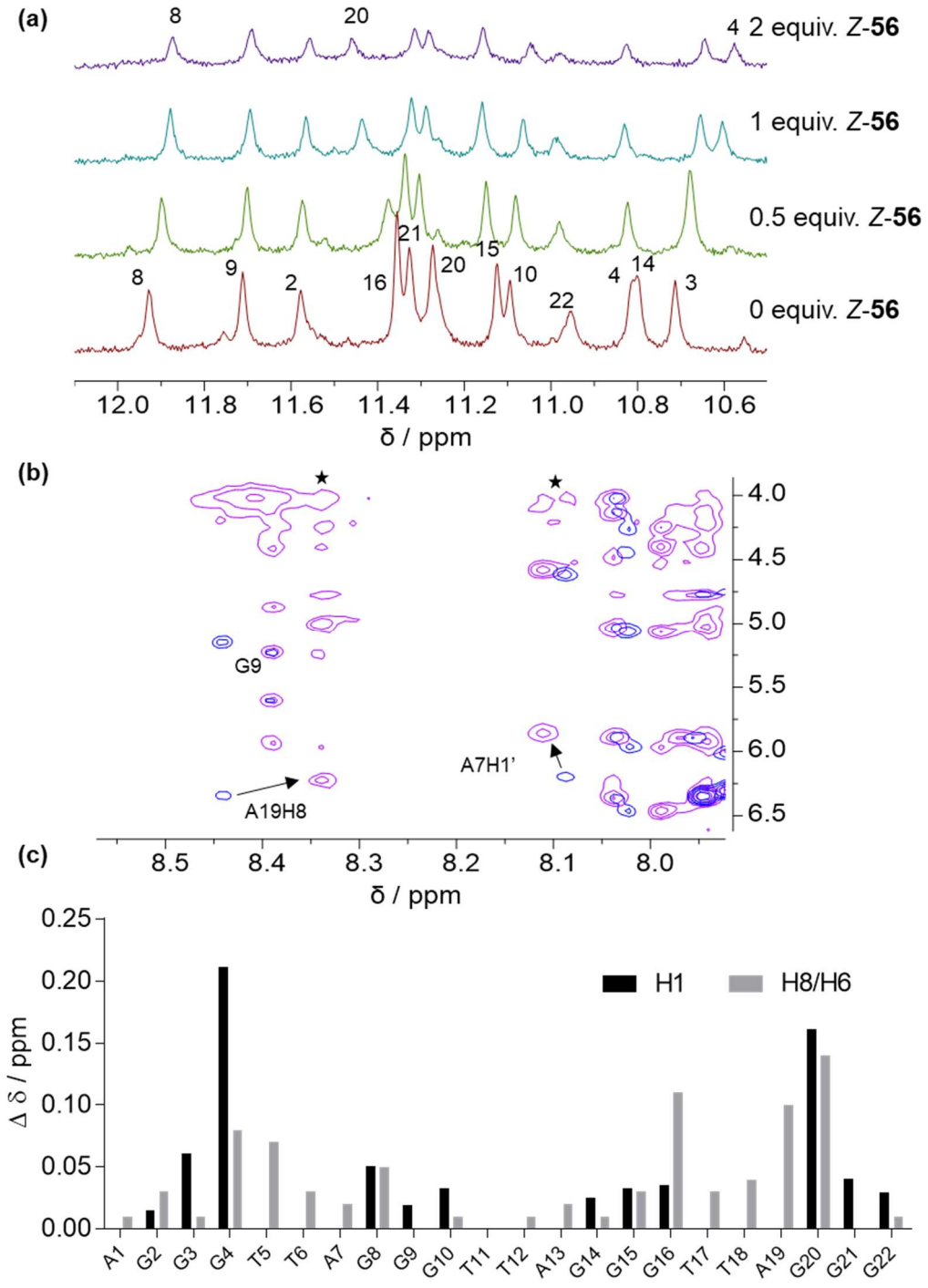


Figure 3.20: NMR invesitgations of telo22-Na^+/*Z*-**56** complex. (a) titration of telo22-Na^+ (imino region) with *Z*-**56**; (b) overlayed NOESY spectra (aromatic/carbohydrate region) of telo22-Na^+ in the absence (blue) and presence (pink) of 2 equiv. *Z*-**56**; (c) chemical shift perturbations of imino and aromatic resonances of telo22-Na^+ upon addition of 2 equiv. *Z*-**56**.

Moreover, an intermolecular NOE correlation is observed between the ligand *N*-methyl group (δ = 4 ppm) and A19H8 and A7H1' protons in the DNA ligand complex (asterisks, Figure 3.20b), providing further evidence for binding to the top fact of the G4 as observed in the computational studies. However, whilst the computational results indicate the ligand to be stacked atop the capping residues of the lateral loops, the fact that G8 and G4 imino resonances are also specifically perturbed suggest stacking interactions with these bases, suggesting that the ligand may open the capping residues to access the top G-tetrad. Meanwhile, resonances corresponding to other regions of the telo22 G4, particularly the lower G-tetrad and diagonal loop are comparatively unperturbed (e.g. G9, as shown in Figure 3.20b), indicating the ligand does not associate with these regions of the G4. These experimental observations support the binding modes identified in the computational studies, which also reveal binding to the top face of the G4, and especially the importance of the A7 and A19 bases in ligand binding.

*3.3.4.4 Binding of ligand E-***57**

Finally, ligand *E*-**57** was investigated using molecular dynamics (Figure 3.21) and NMR (Figure 3.22) studies. In the dynamic simulations, the dihedral distribution (Figure 3.21a) identified three stable conformations. Poses A and B show different orientations of the ligand, but in both cases it appears to bind to the groove of the G4 where it partially interacts with bases from the G-tetrads (G4, G8 and G9), and one of indane residues stacks with the T5 base in the lateral loop. As the simulation progresses however, a further binding pose is observed (Pose C) where the ligand intercalates into the major groove, rupturing the hydrogen bonds in the G-tetrad. In this pose, the ligand is intercalated in between G2, G3 and G15 bases from the middle and lower G-tetrad. Following a further 500 ns simulation, the DNA backbone RMSD (Figure 3.21b) begins to fluctuate between 3.0 to 5.0 Å which suggests a loss of stability of the DNA fold due to rupturing of the G-tetrad hydrogen bonds caused by the intercalation. Ligand *E*-**57** can therefore be expected to induce instability in telo22 G4 DNA, perhaps explaining the attenuation of spectral features observed in the CD titrations (Figure 3.11c). NMR studies confirmed the proposed effect of the ligand more directly, since titration of telo22 with the ligand induced loss of intensity of G4 imino protons and that could be expected to arise from the breakage of the Hoogsteen network. Indeed, at 2 equiv. ligand, the G8 resonance is not visible at all and a new signal in the Watston-Crick region of the NMR spectrum (12.5 ppm) is observed, indicating the ability of ligand partially disrupt the G4 structure, as suggested by the circular dichroism experiments.

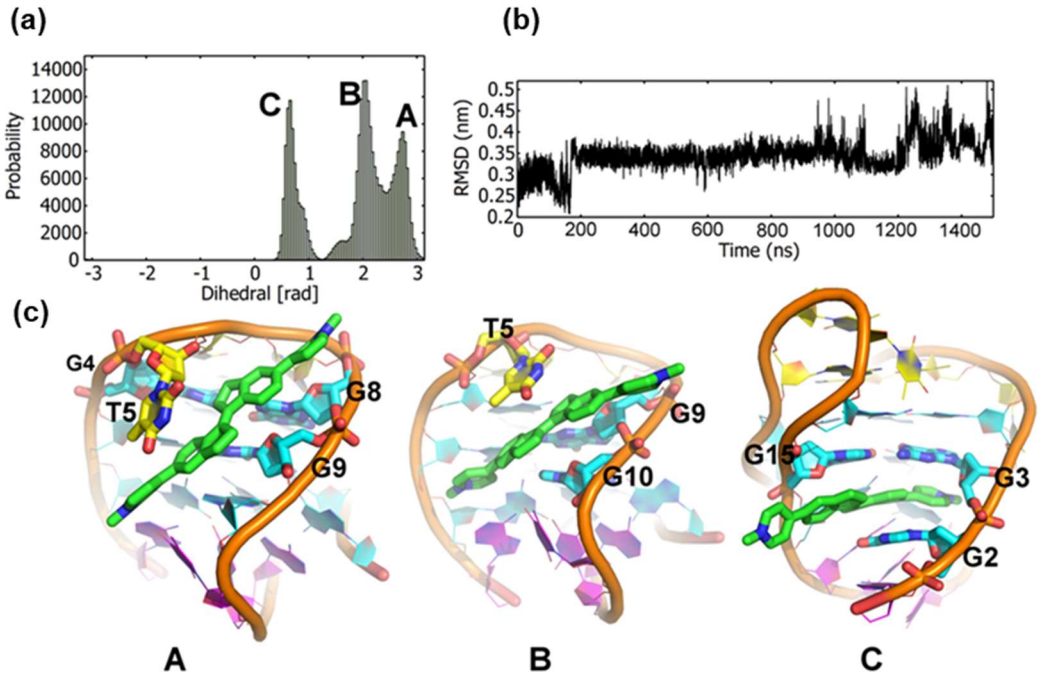


Figure 3.21: Molecular dynamics simulations of telo22-Na^+ and ligand *E*-**57**. (a) probability histogram of ligand orientations sampled, characterised by the dihedral angle between ligand and G4 axes; (b) stability of the G4 backbone co-ordinates over the time course of the simulation (c) binding poses corresponding to maxima A, B and C in Figure (a). *Figures prepared by Dr Susanta Haldar.*

3.3.5 Conclusions of the biophysical studies

From the biophysical assays, it can be concluded that the new *N*-methylpyridinium derivatives interact strongly with different types of G4 DNA and this structural modification offers significant improvement over the more flexible amine-decorated compounds introduced in Chapter 2. The nature of the DNA/ligand interactions depends both on the overall G4 folding conditions and topology, as well as the ligand structure. Towards the aims of this thesis, the different effects of the *E*-**56** and *Z*-**56** ligands on G4-folding in sodium-rich conditions suggest an application for this ligand pair as a photoresponsive chaperone to allow the reversible formation/disruptiuon of G4 under these conditions, as a possible tool in the development of new nanodevices. This possibility is further investigated in Chapter 4.

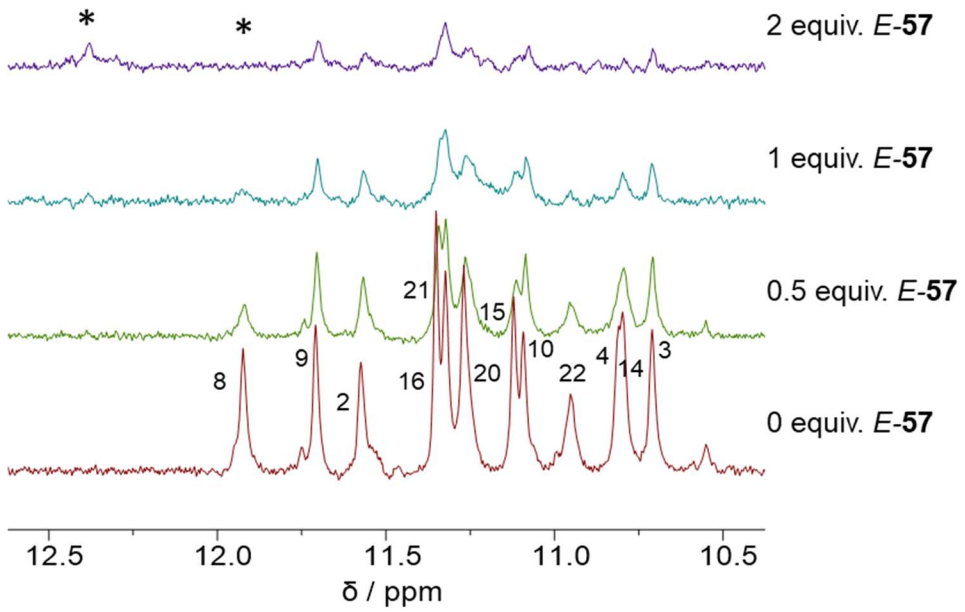


Figure 3.22: NMR titration telo22-Na^+ antiparallel G4 with ligand *E*-**57**. The disappearance of G8 (11.9 ppm) and the appearance of a new signal in the Watson-Crick region (12.4 ppm) are indicated with asterisks. DNA concentration = 185 μM

Meanwhile, towards possible biological applications, it was confirmed that G4 structures were stable in the presence of both *E*-**56** and *Z*-**56** in physiologically-relevant potassium-buffered conditions, as evidenced by CD spectroscopy and NMR studies. However, the more planar *E*-**56** is a significantly more potent ligand, evidenced by FRET thermal melting assays and the UV/visible titrations, whilst NMR and CD studies demonstrated different binding modes for the two compounds. Since the G4 affinity of the more active compound (*E*-**56**) is significantly higher than the amine-decorated compound *E*-**41** reported in Chapter 2, but the selectivity against duplex is retained, the difference in activity of *E*-**56** and *Z*-**56** offered new grounds for optimism that the ligand may ultimately be deployed as a tool for the photoresponsive control of G4 stability *in vivo* as the basis of a photopharmacological strategy. To validate this approach, an initial assessment of the toxicity of the two distinct ligand forms *in cellulo* was undertaken before embarking upon photochemical studies.

3.4 *In vitro* assays

To gain an estimation of the therapeutic potential of the *N*-methylpyridinium G4 ligands, the compounds were screened against a small panel of mammalian cell line and parasitic

organisms. As in Chapter 2, HeLa cervical cancer cells were chosen as a cancerous cell model and MRC5 lung fibroblasts as the non-tumoral model. Compounds were also assessed as antiparasitic agents in *T. brucei* and *L. major* parasites owing to the recent identification of G4 in the genomes of these organisms. Finally, the uptake of a representative compound was studied by confocal microscopy.

3.4.1 Toxicity assays

3.4.1.1 Mammalian cells

Compounds *E*-**56,** *Z*-**56** and *E*-**57** were initially tested for activity in mammalian cells after 72-hour exposure. The dose-response curves are shown in Figure 3.23 and the corresponding EC_{50} values from the fitted curves in Table 3.6. These initial results appeared rather disappointing, with the compounds appearing to be only weakly toxic to HeLa cells.

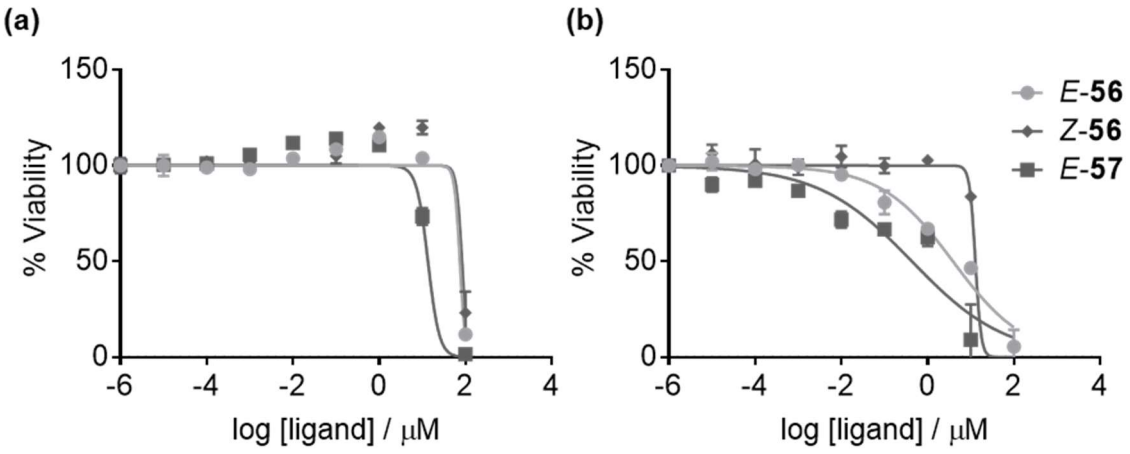


Figure 3.23: Dose-response curves for ligands *E*-**56,** *Z*-**56** and *E*-**57** against (a) HeLa and (b) MRC5 cells at 72 h exposure, measured by the alamarBlue cell viability assay.

Table 3.6: EC_{50} values and selectivity indices for compounds *E*-**56,** *Z*-**56** and *E*-**57** against mammalian cells at 72 h exposure.

Ligand	EC_{50} / μM		Selectivity index
	HeLa	MRC5	(72 h)
E-**56**	77 ± 4	3.1 ± 1	< 1
Z-**56**	82 ± 2	12 ± 1	< 1
E-**57**	16 ± 3	0.35 ± 0.1	< 1

Moreover, the compounds appeared to be more toxic (~ 10-50-fold) to the non-tumoral MRC5 cell-line. This poor discrimination was not initially considered promising given the ultimate desire to deploy these compounds as anticancer agents. However, previous work by others has identified that a longer exposure time to G4-binding compounds is sometimes necessary to observe toxic effects, since a mechanism of action involving telomere shortening theoretically requires several population doublings to take effect.[118,210] For example, BRACO-19 (**7**, Figure 1.14) was demonstrated to show antiproliferative activity consistent with telomere shortening at sub-cytotoxic doses at 15-day exposure.[118] Therefore, the toxicity of the compounds after longer (7 day) exposure were measured (Figure 3.24 and Table 3.7). Strikingly, whilst the toxicity towards MRC-5 was comparable with that observed at shorter-term exposure (EC_{50} = 10-100 μM), the cancerous cells became more susceptible to lower doses of compound *E*-**56** (EC_{50} = 62 nM), the strongest G4 ligand in the compound series. This increase in potency returns a selectivity index of 29 between the cell lines at long-term exposure. That the toxicity to MRC-5 cells does not significantly depend on ligand exposure time for could be attributed to the telomerase-negative nature of this cell line, rendering this model less susceptible to effects of telomeric G4 binding ligands. The comparable toxicity observed for MRC5 at 3-and 7-day exposure may therefore result from an alternative mechanism of action not related to telomere shortening.

Meanwhile, whilst the toxicity of the comparatively *Z*-**56** is slightly increased at longer exposure times, the effect is much less pronounced than for compound *E*-**56**. The longer exposure increases the activity of *E*-**56** by three orders of magnitude, whilst the potency of *Z*-**56** increases by only one order of magnitude. This difference in effect correlates with the G4 binding properties of the ligands. The net result is that at 7-day exposure, *E*-**56** appears approximately 85 times more potent (EC_{50} = 62 nm) than the *Z*-**56** isomer (EC_{50} = 5.2 μM). This result is particularly relevant in light of the goal to deploy the ligand as a photopharmacological agent, since it suggests that the biological activity of the G4-interactive compounds may ultimately be controlled by photoisomerisation.

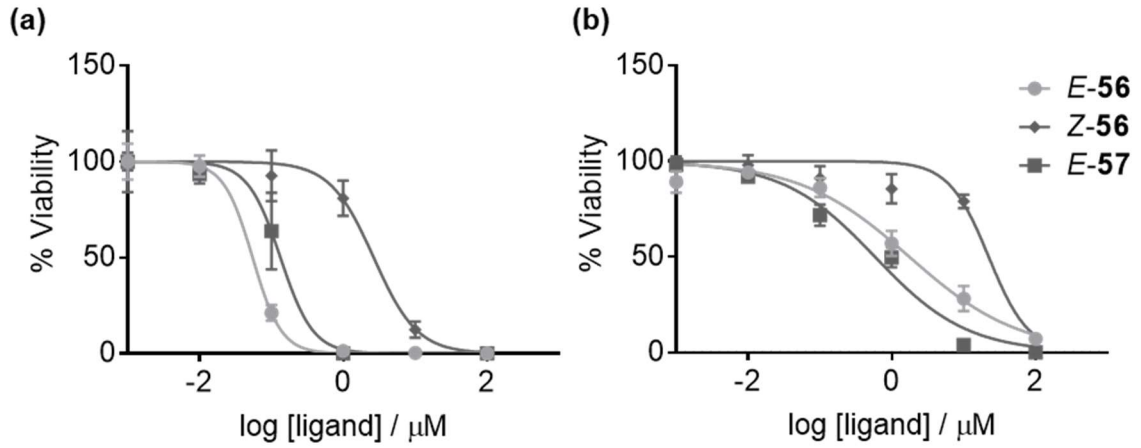


Figure 3.24: Dose-response curves for ligands *E*-**56,** *Z*-**56** and *E*-**57** against (a) HeLa and (b) MRC5 cells at 7d exposure, measured by the alamarBlue cell viability assay.

Table 3.7: EC_{50} values and selectivity indices for compounds *E*-**56**, *Z*-**56** and *E*-**57** against mammalian cells at 7-day exposure.

Ligand	EC_{50} / μM		Selectivity index
	HeLa	MRC5	(7 day)
E-**56**	0.062 ± 0.01	1.8 ± 0.5	29
Z-**56**	5.2 ± 2	20 ± 5	3.8
E-**57**	0.11 ± 0.3	0.46 ± 0.1	4.2

3.4.1.2 Parasitic organisms

Next, efforts turned to considering the potential of the G4 ligands as antiparasitic agents. Activity was first measured against both *T. brucei* and *L. major,* with MRC-5 cell lines chosen as non-tumoral cell model for comparison (Figure 3.25 and Table 3.8). Pleasingly, compounds *E*-**56,** *Z*-**56** and *E*-**57** exhibited a potent toxicity to *T. brucei* with EC_{50} values in the nanomolar range. Critically, the compounds are much less toxic against the non-tumoral MRC-5 cells with selectivity index of up to 280-fold, suggesting a promising therapeutic window for these compounds as anti-trypasonomal agents. Though the compounds appear to exert a lower toxicity to the *L. major* parasite, ligands *E*-**56** and *E*-**57** show sub-micromolar efficacy against this organism, suggesting a potential role for G4 recognition in their mechanism of action in this case. The selectivity index is significantly lower than observed in the case of *T. brucei,* but

the values are comparable to those observed for other G4 ligands screened against this organism.[52]

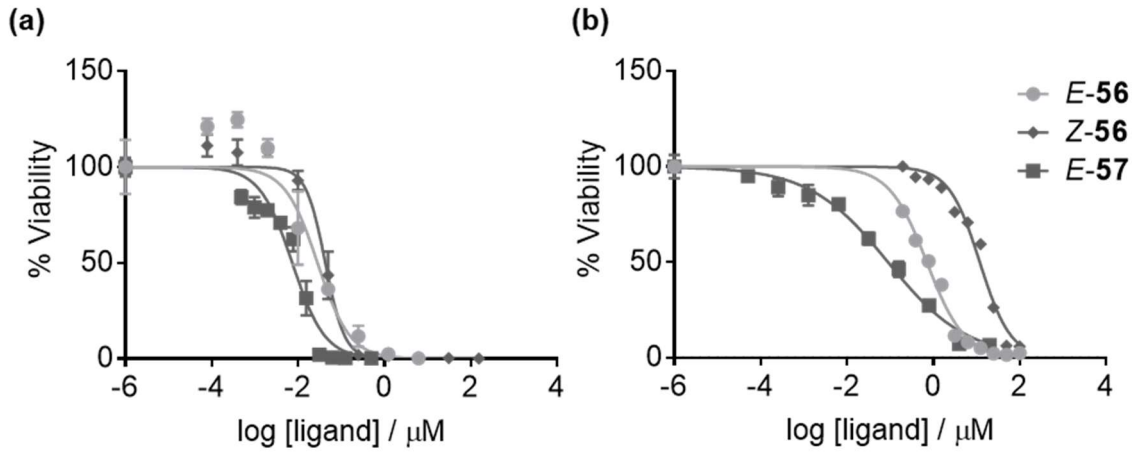


Figure 3.25. Dose-response curves for ligands *E*-**56,** *Z*-**56** and *E*-**57** against (a) *T. brucei* and (b) *L. major* parasites at 3d exposure, measured by the alamarBlue cell viability assay (*T. brucei*) or MTT assay (*L. major*).

Table 3.8: EC_{50} values and selectivity indices for compounds *E*-**56,** *Z*-**56** and *E*-**57** against parasitic organisms at 72 h exposure.

	EC_{50} / μM		Selectivity index (72 h)	
Ligand	*T.brucei*	*L.major*	*T.brucei*	*L.major*
E-**56**	0.0029 ± 0.0005	0.64 ± 0.1	110	4.8
Z-**56**	0.043 ± 0.0001	11 ± 2	280	1.1
E-**57**	0.0061 ± 0.002	0.11 ± 0.03	60	3.2

3.4.2 Confocal microscopy

As a step towards investigating the mechanism of action of ligands *E*-**56,** *Z*-**56** and *E*-**57** the uptake and localisation of the compounds by mammalian cells and parasites was examined through microscopy studies. Unfortunately, lead compound *E*-**56** was not sufficiently fluorescent to visualise using this technique. Studies therefore focused on compound *E*-**57**, which has photophysical properties (λ_{em} = 550 nm) much better suited to visualisation in cells (Figure, 3.26). After 30 min incubation at 37 °C, significant uptake of compound *E*-**57** in both HeLa cells (Figure 3.27) and *T. brucei* parasites (Figure 3.28) was observed. The ligand *E*-**57** was mainly localised in the nucleolus and the cytoplasm of HeLa cells. In *T. Brucei*, the same

ligand was mainly found in the nucleus and the kinetoplast. These results suggest that pyridinium stiff-stilbene G4-ligands can reach DNA harbouring sites and therefore possibly target G4 structures. Indeed, G4 DNA structures have recently been directly observed in the nucleoli of live cells.[211] Further investigations into the therapeutic mechanism of stiff-stilbene ligands is warranted in order to interrogate the true biological targets and obtain more conclusive evidence regarding the potential G4-mediated mode of activity.

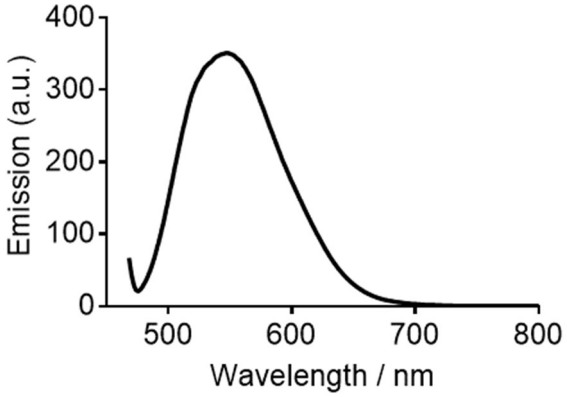


Figure 3.26: Emission spectrum of ligand *E*-**57** in water (0.5% DMSO). λ_{ex} = 450 nm. Ligand concentration = 50 μM.

3.5 Chapter conclusions

The work reported in this chapter built upon the results of Chapter 2 and further develops the stiff-stilbene ligand core into a system more suited to biological applications through optimisation of the G4 binding properties. These studies led to the identification of *N*-methylpyridinium stiff-stilbene compounds as potent G4-binding agents, demonstrating significant improvements over the first-generation amine-based ligands. In particular, the new compounds exhibit higher affinity for G4 structures, demonstrated through a variety of biophysical assays, an effect which appears to translate into biological effects that correlate to some degree with the G4 binding properties, though the mechanism of action remains to be rigorously elicited. Towards the overall goal of developing a photoresponsive G4 binding agent as the basis of a molecular fuel or photopharmacology, the very different activities of the *E*-**56** and *Z*-**56** ligand pair, both in terms of toxicity and of their effect on G4 folding topologies were particularly significant, suggesting the potential to control ligand activity through photoisomerisation. Efforts therefore turned to attempts to investigate the photochemistry of the *E/Z*-**56** pair towards this end, as reported in the following chapter.

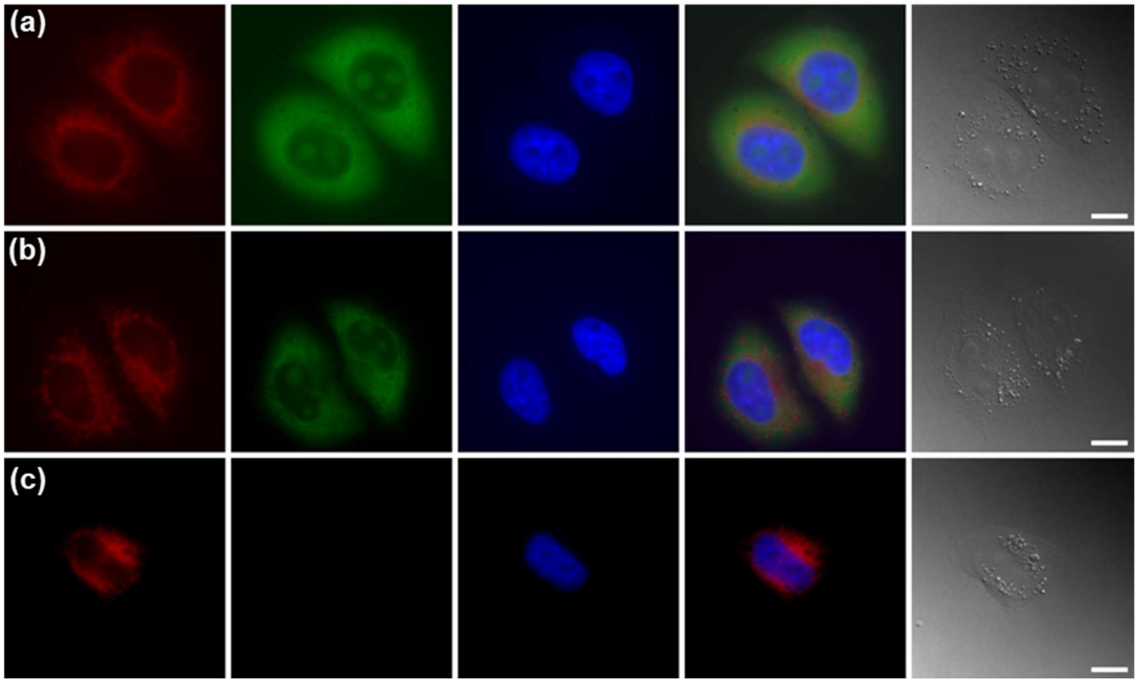


Figure 3.27: Fluorescence microscopy images of HeLa cells after incubation with 0.5 μM ligand *E*-**57** for (a) 30 min or (b) 120 min. Control (c) cells without added compound are also displayed. Visualisation panes from left to right: mitochondria (Mitotracker Red), ligand *E*-**57**, nucleus (DAPI), co-localisation image, bright-field image. Scale bar: 20 μm.

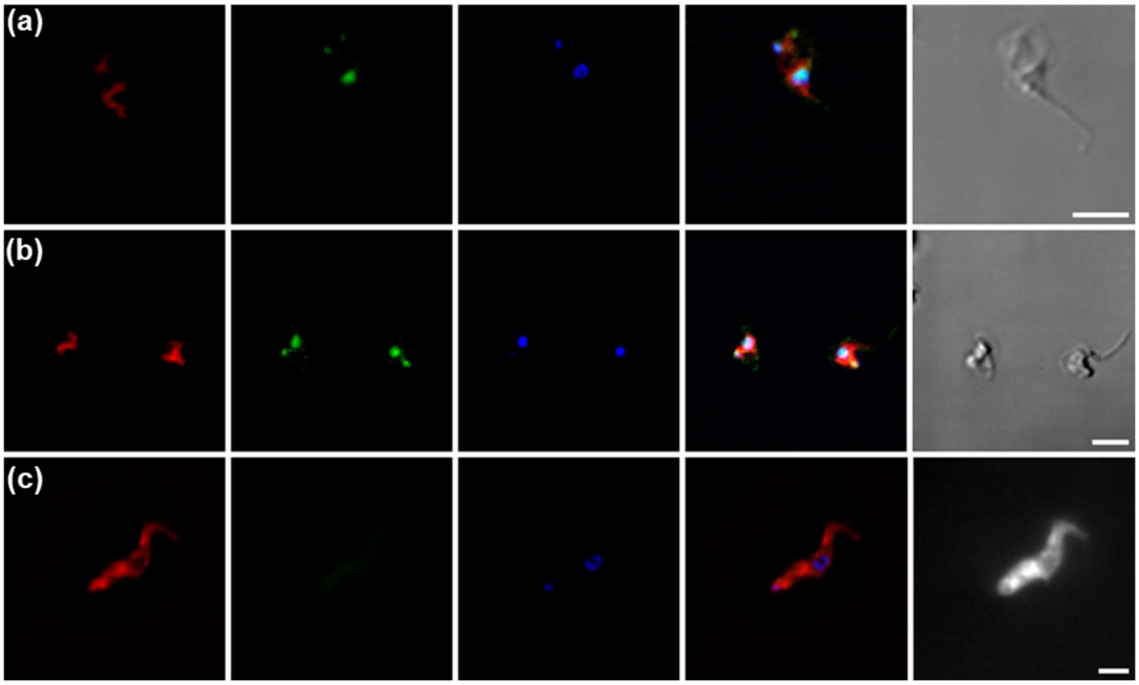


Figure 3.28: Fluorescence microscopy images of *T.brucei* parasites after incubation with 1 μM ligand *E*-**57** for (a) 30 min or (b) 120 min. Control (c) parasites without added compound are also displayed. Visualisation panes from left to right: mitochondria (Mitotracker Red), ligand *E*-**57**, nucleus (DAPI), co-localisation image, bright-field image. Scale bar: 5 μm.

4 G4 photocontrol with stiff-stilbene ligands

This chapter is based upon work published in *Angewandte Chemie: International Edition* of which I am first author.[203] I was responsible for all elements of experiment design and data collection. All experiments were all undertaken at the University of Bristol, UK. Assistance with sample irradiation was provided by Dr Thomas A.A. Oliver and Miss Marta Duchi.

4.1 Introduction

The overarching goal of this project was the design of a new class of photoresponsive G4 ligand which allowed activity to be controlled by light in physiologically-relevant conditions. Towards this end, the previous chapters report the development of a set of ligands designed from the stiff-stilbene photoswitch, indicating that the G4 binding properties, effect on G4 structure and cytotoxicity may be determined by the configuration of the central stilbene double bond of the *E/Z*-**56** ligand pair. Given that stiff-stilbenes have previously been demonstrated to undergo reversible *E*/*Z* isomerisation upon photoirradiation,[172,178] the *E/Z*-**56** ligand couple looked very promising as a tool to allow G4 regulation by light with a ligand-driven strategy, as summarised in Figure 4.1. In the current chapter, investigations of the photochemistry of the *E/Z*-**56** ligand couple are reported, and the observed photoresponse is used as a tool to photoregulate G4 DNA structure.

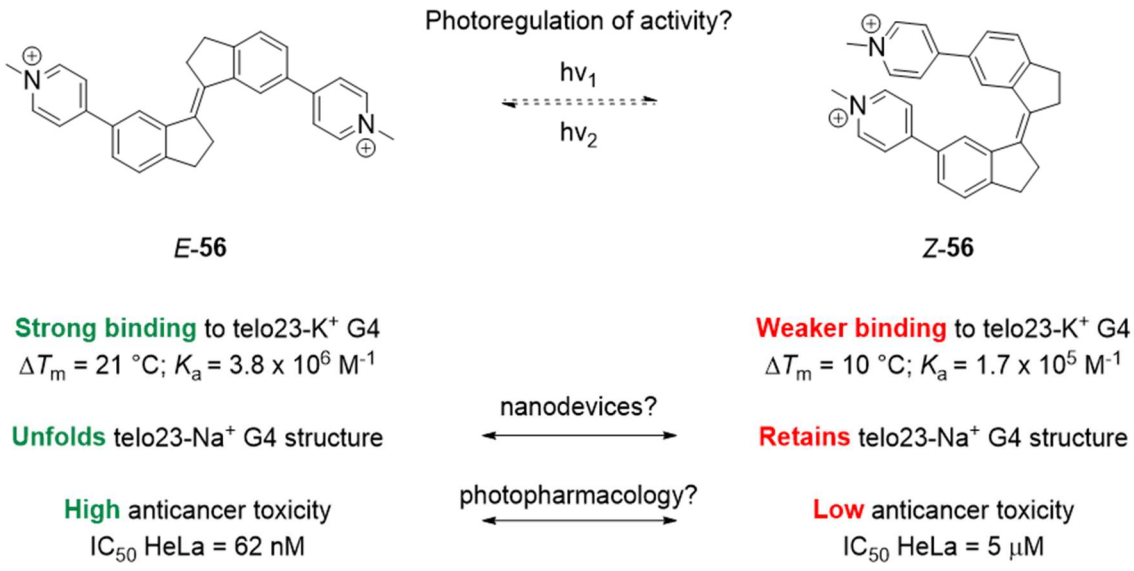


Figure 4.1: Proposed concept of photoresponsive G4 ligand control.

4.2 Ligand photochemistry

As a starting point towards applying the *E/Z*-**56** pair as a tool for the photoregulation of G4 DNA, the photochemistry of the ligand was first characterised in organic and aqueous environments. These studies were required in order to developing an understanding the behaviour of the ligand upon photoirradiation that could ultimately be deployed to regulate the G4 DNA structure.

4.2.1 Photochemistry in DMSO

The photochemistry of ligand *Z*-**56** was initially investigated in DMSO-d_6. Though this solvent is ultimately unsuitable for the DNA-based applications of the photoresponsive ligand, the ligand was significantly more soluble in this medium allowing higher concentrations to be achieved, meaning the photochemistry could be conveniently monitored *via* ^{1}H NMR spectroscopy. 400 nm was selected as the irradiation wavelength since a bright monochromatic light source was available at this wavelength and it is towards the blue-shifted end of the absorbance of *E*-**56** and *Z*-**56** (see UV/visible spectra in Section 3.3.2) and therefore more suited to biological application, since high-energy UV light is deleterious to cells. Upon irradiation of a 1 mM sample of *Z*-**56** in DMSO-d_6, photoisomerisation to the *E*-**56** form is clearly observed in the ^{1}H NMR spectra (Figure 4.2). The isomerisation appears to proceed cleanly, since all new resonances can be assigned to the formation of the *E* isomer, and no by-products are detected. The conversion was monitored at intervals up to 90 minutes, by which time no further conversion was detected indicating the 400 nm photostationary state (PSS) had been achieved. This PSS consists of an approximately 1:1 mixture of *E* and *Z* ligand forms. Though this does not represent complete conversion, that the *E*-**56** isomer is 85-fold more toxic than the *Z*-**56** form in the cellular assays (see Section 3.4) suggests that even this partial conversion could result in suitable *in situ* activation of the molecule if delivered at a suitable concentration. However, though this initial study was promising, such an effect would need to be achieved in aqueous cell-like media to be of clinical utility. Efforts therefore turned to investigating the behaviour of the ligand in buffered aqueous solutions.

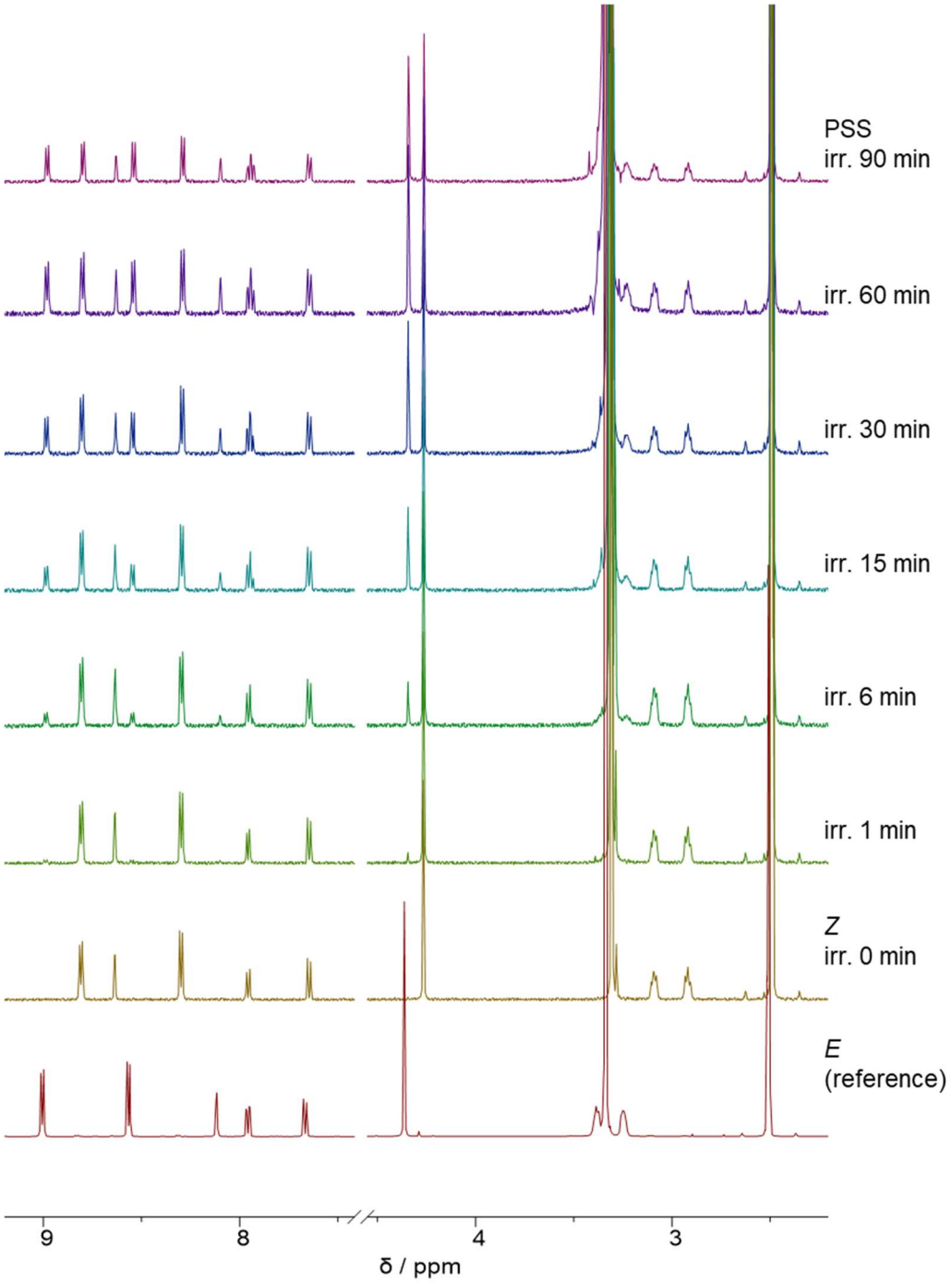


Figure 4.2. Photogeneration of ligand *E*-**56** by irradiation of *Z*-**56** with 400 nm light, monitored by ^{1}H NMR spectroscopy in DMSO-d_6. [Ligand] = 1 mM.

4.2.2 Photochemistry in buffered aqueous solution

In order to understand the ligand photochemistry in the aqueous conditions relevant to G4 folding, photoirradiation of *Z*-**56** was performed at 10 μM concentration in 100 mM sodium phosphate buffer, at which concentrations both forms of the ligand are water-soluble and the photoreaction can be monitored by UV-absorbance spectroscopy (Figure 4.3). Surprisingly, when a sample of *Z*-**56** was irradiated under these conditions, the UV spectra instead indicate the conversion of the ligand to a new photoproduct rather than isomerisation to *E*-**56** (Figure 4.3a). Specifically, the absence of isosbestic points (which would be expected for direct conversion between two species without the formation of a stable intermediate) and the attenuation of absorption at λ > 325 nm indicate an alternative reaction pathway. The same phenomenon was observed both in pure H_2O (Figure 4.3d) and irradiation of the *E*-**56** isomer under the same conditions generated the same UV/visible signature (Figures 4.3b and 4.3e), indicating that both ligand isomers ultimately form the same photoproducts. It should be noted that in control experiments, the stability of the ligand in solution in the dark proved that the reaction pathway is light-mediated rather than a result of inherent instability of the ligand in aqueous conditions (Appendix A4, Figure A4.1).

To determine the fate of ligand *Z*-**56** under the reaction conditions, the photoproduct was analysed by mass spectrometry and NMR spectroscopy. Mass spectrometry analysis (Figure 4.4) indicated evidence of the oxidative decomposition of the ligand. In particular, a signal was observed for a species with *m/z* = 224.1 which corresponds to both ketone **61** and dioxirane **60** (Figure 4.5). A signal corresponding to a molecular ion *m/z* = 210.1, consistent with demethylation of the pyridinium moiety to form product **62** was also observed.

Formation of a ketone was clearly visible in the NMR spectra of the photoproduct (Figure 4.6). In particular, the presence of a carbonyl group is indicated in the ^{13}C NMR spectrum, δ = 206 ppm (Figure 4.6b). Surprisingly, the NMR spectrum also indicated the loss of the *N*-methyl group, though trace amounts of *N*-methylated products were visible in the ^{1}H NMR spectrum (δ = 4.25 ppm). Owing to the difficulty of performing this reaction on a practical scale due to the low solubility of the ligand under the reaction conditions, efforts to purify and characterise these minor by-products were not pursued.

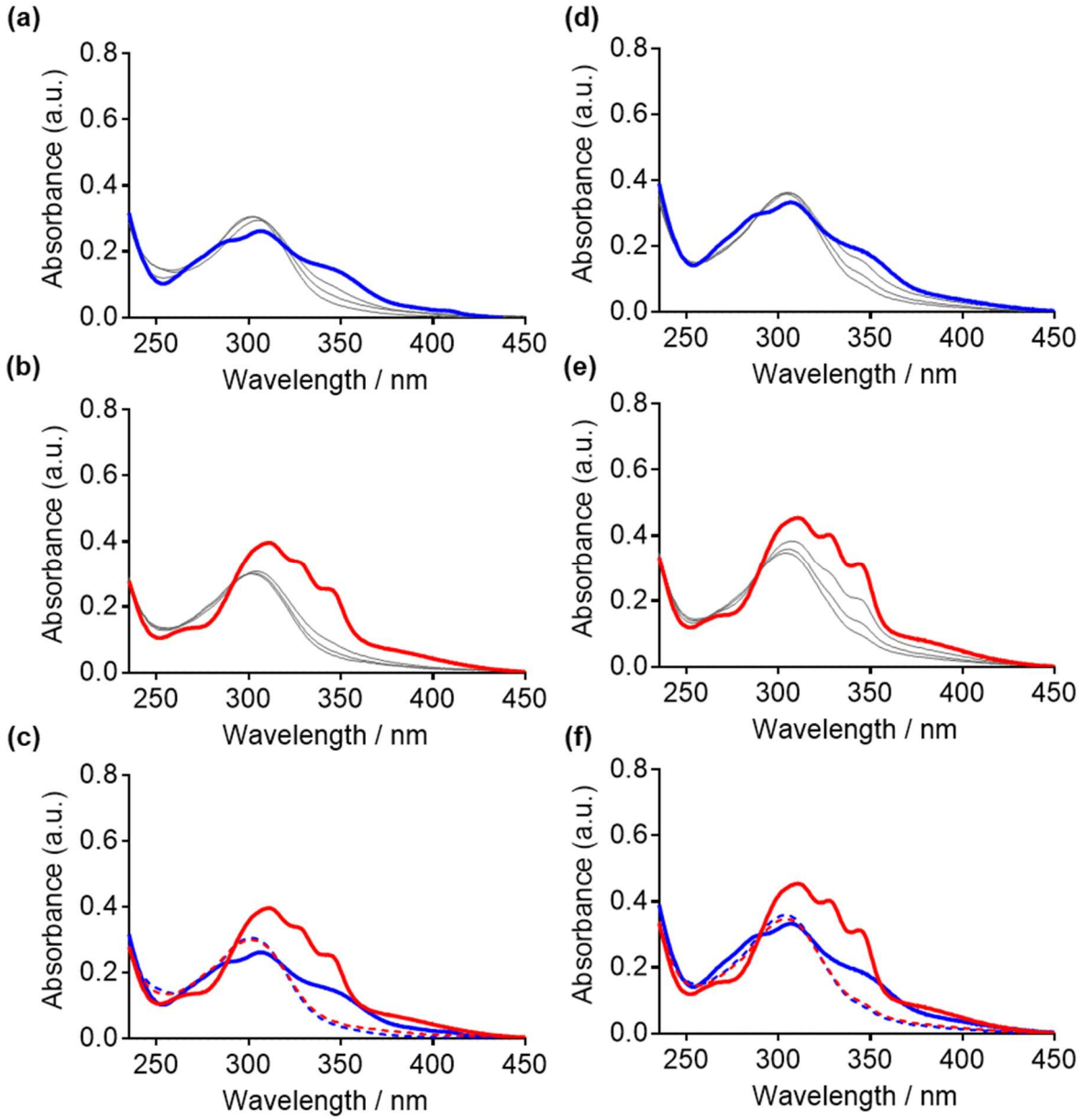


Figure 4.3. Photoresponse of ligands *E*-**56** and *Z*-**56** (10 μM) irradiated at 400 nm in (a-c) 100 mM sodium phosphate buffer, pH 7.4 and (d-f) pure H_2O. (a),(d) irradiation of *Z*-**56** (solid blue trace) for 30, 60 and 90 min (grey traces); (b),(e) irradiation of *E*-**56** (solid red trace) for 30, 60 and 90 min (grey traces); (c),(f) comparison of spectra of *E*-**56** and *Z*-**56** (solid traces) and the spectra of the respective photoproduct (dotted traces) after 90 min irradiation indicating the formation of the same species from both ligands.

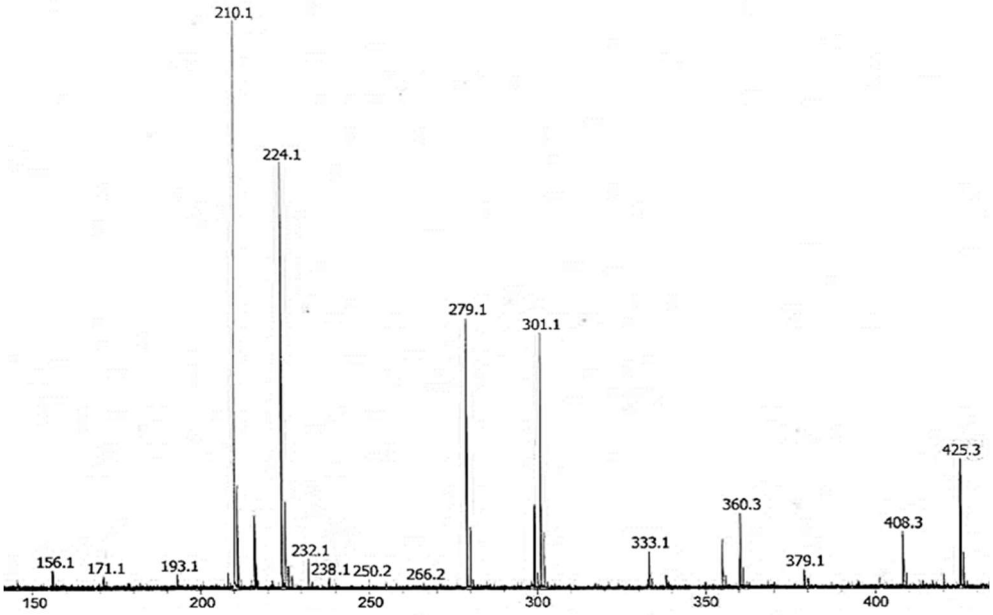
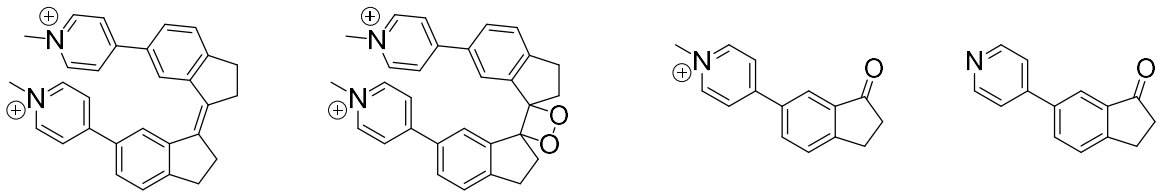


Figure 4.4: Mass spectrometry analysis of the photoproduct after irradiation of *Z*-**56** in 100 mM sodium phosphate buffer.



m/z: 208.1 $[M]^{2+}$ *Z*-**56**

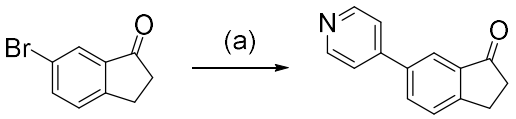
m/z: 224.1 $[M]^{2+}$ **60**

m/z = 224.1 $[M]^{+}$ **61**

m/z: 210.1 $[MH]^{+}$ **62**

Figure 4.5: m/z values of possible photoproducts of ligand *Z*-**56**

In order to confirm the identity of the major product, a pure sample of this material was synthesised by Suzuki coupling of bromide **42** with 4-pyridylboronic acid (Scheme 4.1). The spectral data of this reference compound were in agreement with the photoproduct (Figure 4.6).



42 **62**, 28%

Scheme 4.1: Synthesis of ketone **62**. Reagents and conditions: (a) 4-pyridylboronic acid, $Pd(PPh_3)_4$ (20 mol %), Na_2CO_3, PhMe/EtOH (9:1), 100 °C, 17 h.

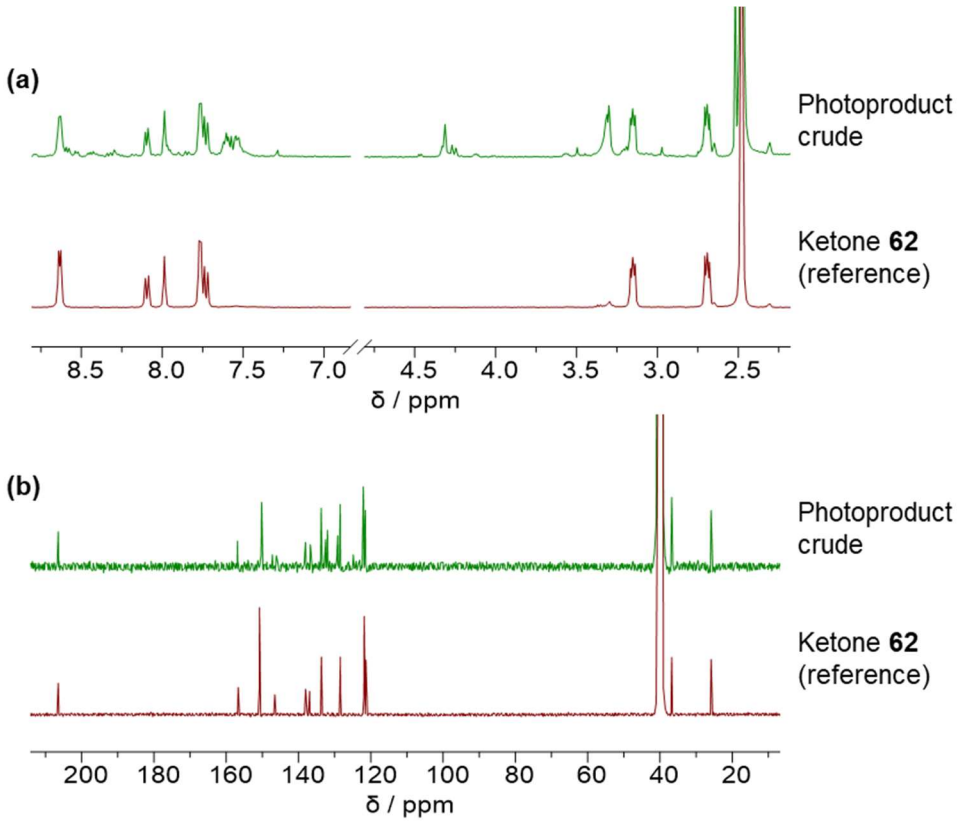
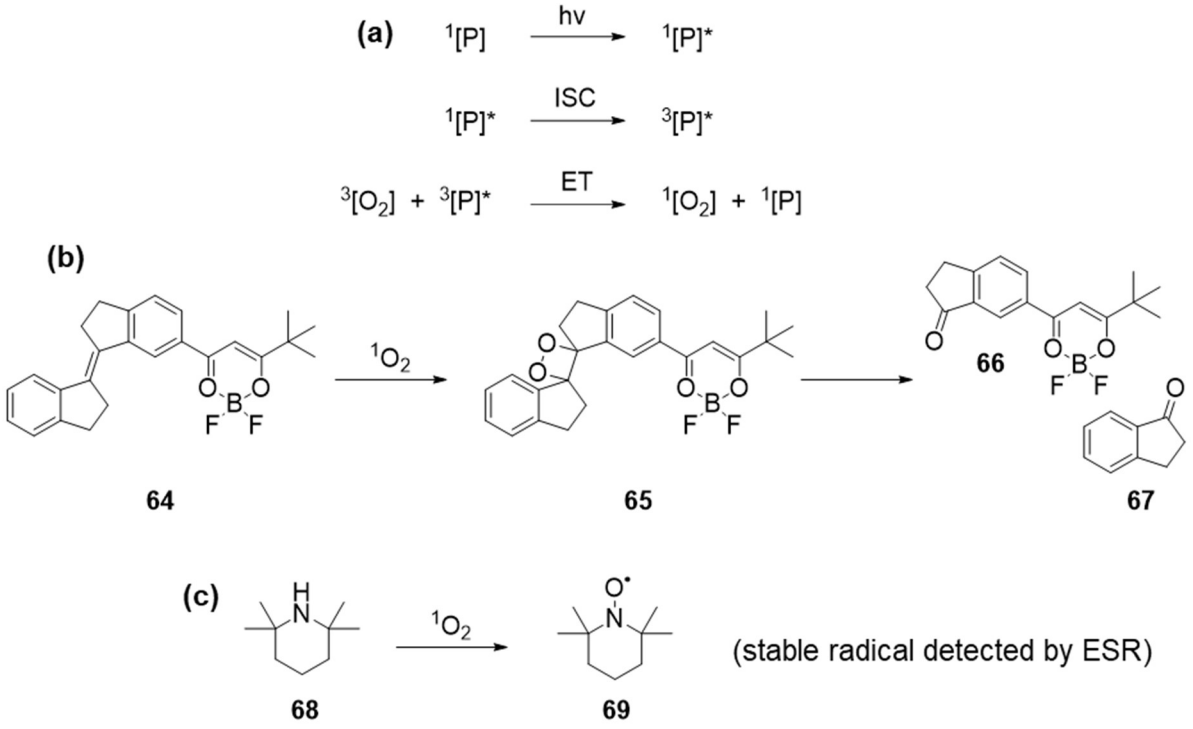


Figure 4.6: NMR analysis of the *Z*-**56** photoproduct and comparison with a reference sample of ketone **62**. (a) ^{1}H NMR spectra; (b) ^{13}C NMR spectra.

4.2.3 Mechanistic comments

In-depth mechanistic studies of the photoreaction of ligand *Z*-**56** were not undertaken during the course of this project. However, the observations discussed above can be considered in the context of precedent reported in the literature to rationalise the observed reactivity.

Reaction with ground state triplet oxygen is forbidden on spin conservation grounds, and therefore the ligand must act as a photosensitiser to generate the reactive species if this mechanism is to be possible.[212] Addition of singlet oxygen across the stilbene olefin of *E/Z*-**56** generates *endo*-peroxide **60** and subsequent fragmentation yields two equivalents of ketone **61**. Loss of the methyl group is surprising, but demethylation of related systems has previously been observed, albeit under basic conditions;[213] possibly this moiety becomes more labile during the photoreaction. Subsequent to the publication of this work, a similar self-sensitised reaction pathway was also proposed by Boulatov and co-workers for a related stiff-stilbene system, **64**, where the dependency of the reaction on oxygen concentration was ultimately employed in the development of a sensor for intracellular oxygen levels (Scheme 4.2).[214] In this case, the authors observed the generation of singlet oxygen by using scavenging experiments with TEMP **68** to generate TEMPO **69**, which may be detected by electron spin resonance spectroscopy (Scheme 4.2c).



Scheme 4.2: Photosensitised pathway for oxidation of stiff-stilbenes recently proposed by Boulatov *et al.* towards the generation of an oxygen sensor (a) mechanism of self-sensitised generation of singlet oxygen by intersystem crossing and energy transfer from a photosensitiser (P); (b) oxidation and fragmentation to form the ketone products; (c) use of TEMP to detect formation singlet oxygen.

4.3 Phototriggered control of G4 folding

Though the eventual fate of ligands *E*-**56** and *Z*-**56** upon photoirradiation in water was a disappointment regarding the ultimate application of the ligand pair as a tool for reversible *in situ* control of G4, the observed photoreactive pathway was nonetheless considered a possible means to fuel repeated switched in G4 conformation by employing the ligand as a photoresponsive fuel.

To examine this hypothesis, the telo23-Na^+ structure was selected since the unfolding effect of ligand *E*-**56** is most effective under these conditions, as previously demonstrated by CD spectroscopy (Section 3.3.3) and NMR spectroscopy (Section 3.3.4). The ability of *E*-**56** to fuel repeated conformational switches in G4 DNA was demonstrated (Figure 4.7). First, the CD was recorded in the absence of ligand (Figure 4.7a, green trace) demonstrating the characteristic signature of the hybrid G4 fold. Addition of 10 equiv. *E*-**56** induced the conformational switch to the unfolded form of the DNA (Figure 4.7a, blue trace) as previously investigated by the CD, NMR and metadynamics simulations. Pleasingly, the CD spectrum acquired following irradiation of the sample with 400 nm light exactly resembled the original spectrum of the G4 in the absence of the ligand (Figure 4.7a, red trace). Therefore, exposure to blue light deactivates ligand *E*-**56** and causes the regeneration of the original G4 folded topology. To examine the reversibility of this process, a further aliquot of *E*-**56** was added to the same DNA solution and the CD spectrum of ligand-bound unfolded telo23-Na^+ was restored almost completely (Figure 4.7a, blue trace). This fuelling/irradiation cycle was repeated several times, and the conformation of the G4 was switched between folded and unfolded states at least five times before significant fatigue of the system was observed (Figure 4.7b). Nonetheless, further partial reversibility continued to be observed to some extent up to 8 times (Figure 4.7c). Further evidence for the reversible unfolding of G4 was obtained by NMR analysis (Figure 4.7d), which show the recovery and sharpening of the imino signals upon photoirradiation, indicating the refolding of the DNA to regenerate the original G4.

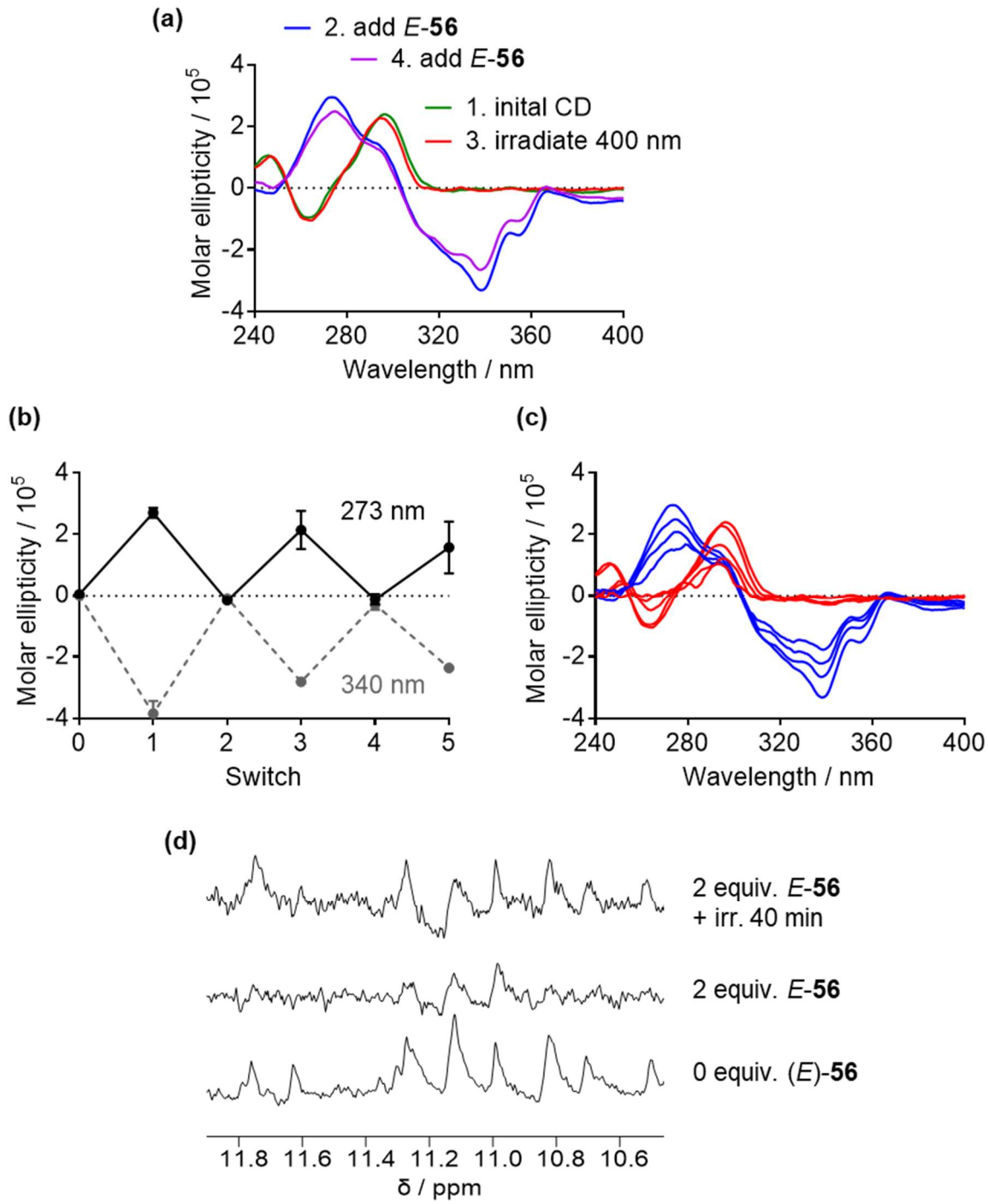


Figure 4.7: Reversible switching of telo23-Na^+ G4 folding/unfolding by ligand *E*-**56** (a) spectra of telo23-Na^+ generated by exposure to 400 nm light in the presence of *E*-**56**; (b) Plot of the change in ellipticity at 273 nm and 340 nm over several dose/irradiation cycles (c) superimposed CD spectra of eight switches showing the eventual photofatigue of the system; (d) unfolding/refolding of telo23-Na^+ monitored by ^{1}H NMR spectrsocopy.

This reversible folding/unfolding of G4 could be considered as a primitive nanodevice, where the photoresponsive ligand induces the metastable unfolded state in the DNA system allowing the G4 structure formation to be triggered on-demand (Figure 4.8).

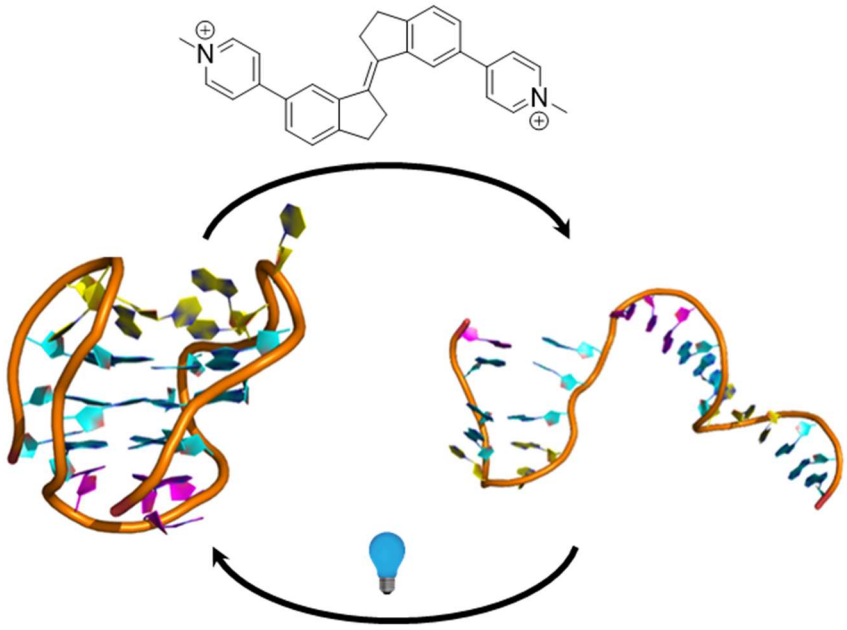


Figure 4.8: Reversible switching of telo23-Na^+ G4 folding/unfolding by ligand.

4.4 Chapter conclusions

In this chapter, the photochemistry of the *E/Z*-**56** ligand couple was investigated with the aim of moderating the G4 binding activity *in situ* towards the development of a G4-targeted photopharmacology or as means to control G4 folding for the development of oligonucleotide-based nanodevices. Unfortunately, the present system is not suitable for photopharmacological applications, since the inactive form of the ligand cannot be converted to the active form in aqueous conditions, instead undergoing irreversible oxidative cleavage. Nonetheless, the ligand permits the photoregulation of G4 folding in the presence of sodium ions without the requirement to pre-incorporate photoresponsive functionality into the biomolecule, suggesting potential applications in the control of natural G4 systems. The approach is complementary to the azobenzene system previously reported,[143] in that the ligand is able to exert dramatic effects on G4 folding in the presence of metal salts and induces unfolding against the cation-stabilised structure, rather than templating G4 folding in the absence of cations. It would however be much more desirable for the system to function without the need for repeated additions of chemical fuels and the generation of waste

products. Efforts therefore turned to the identification of an alternative chromophore as the basis of a fully photoreversible G4 ligand that functions in physiologically-relevant conditions. The results of these endeavours are reported in the next chapter.

5 Dithienylethene G4 ligands

This chapter is based upon work published in *Chemical Communications*.[215] I was responsible for all elements of compound design and synthesis for compounds **77o**/**77c** and **79o**, all photochemistry and biophysical assays, and the writing of the manuscript. The synthesis of compound **78o** was performed by Dr Javier Ramos-Soriano. The cytotoxicity studies of **77o**/**77c** were performed by Miss Sadiyah Sheikh. All experiments were undertaken at the University of Bristol, UK.

5.1 Introduction

The previous chapters reported progress towards the design of stiff-stilbene based photoresponsive G4 ligands. This scaffold proved highly effective as the basis of new G4-targeting molecules, achieving high binding affinity and selectivity for G4 over duplex DNA in the best cases. Critically, the olefin geometry of the stilbene core proved a crucial determinant of the G4 binding activity and biological activity, particularly in the *N*-methylpyridinium compounds *E*-**56** and *Z*-**56**. Unfortunately, in the derivatives examined thus far, photooxidative cleavage of the stilbene core (rather than the desired photoisomerisation) was found to dominate and hence the current stiff-stilbene system cannot yet be deployed towards the proposed biological applications. A ligand to regulate G4 in physiologically-relevant conditions that could be reversibly switched with light, without the generation of by-products, was still sought. Towards this end, this chapter describes the work undertaken to investigate the potential of dithienylethene (DTE) derivatives as supramolecular agents for the photoregulation of G4 structures. The molecular photoswitch is briefly introduced and previous attempts to deploy these derivatives to target DNA-based systems are outlined. The synthesis of a small set of DTE derivates is then reported, and studies of the photochemistry of these derivatives in the absence and presence of G4 and duplex DNA are described. The G4 binding properties of the ligands were investigated using FRET thermal melting assays, UV/visible absorbance spectroscopy, circular dichroism, and NMR spectroscopy. These studies demonstrated that the bistable states of the photoswitchable ligands associate to G4 with different binding modes. Finally, the cytotoxicity of the two states of the photoswitchable ligand were investigated, and the results suggest potential for these ligands to be deployed in photopharmacological strategies.

5.2 Ligand design rationale

5.2.1 Dithienylethene as a bistable photoswitch

Dithienylethenes are a subset of diarylethene photoswitches, the first of which was reported by Irie and Mohri in 1988.[216,217] The design of these switches evolved from the observation that a common side-reaction in the photoisomerisation of stilbenes *E/Z*-**21** is photocyclisation to the dihydrophenanthrene derivative (Figure 5.1a). Whilst barrier height for the thermal disrotatory ring closure in the ground state prevents this reaction from occurring at ambient temperature, the energy barrier is significantly lower in the photoexcited state. In this case, the ring closure proceeds in a conrotatory fashion (on orbital symmetry grounds) to form *trans*-dihydrophenanthrene **70**. Under ambient conditions, rapid oxidation of this species occurs to generate phenanthrene **71** as the final product, rendering this reaction pathway irreversible and therefore unsuitable as the basis of a reversible molecular switch.

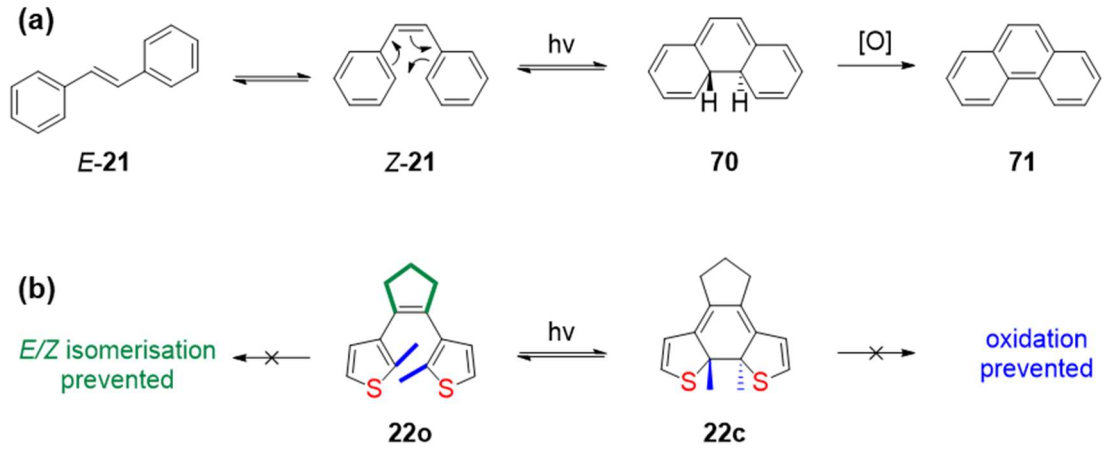


Figure 5.1: Development of dithienylethene photoswitches. (a) limitations to application of stilbene photoswitches: photocyclisation competes with *E/Z* isomerisation which is irreversible owing to oxidation of the dihydrophenanthrene **70**; (b) design features of dithieylethenes as molecular photoswitches to circumvent the limitations of stilbene-derived systems.

This limitation is circumvented by substitution of the *ortho* positions of the aromatic ring with methyl (or other alkyl) groups to prevent oxidation (Figure 5.1b). However, the lifetime of the cyclised form remains short in the stilbene system, since the aromaticity of the open isomer renders this form more stable, resulting in a low barrier height to cycloreversion. To address

this issue, the benzene rings are replaced with heterocycles of lower aromaticity, such as furans or thiophenes. This modification renders the open and closed isomers closer in energy and increases the energy barrier of the thermal disrotatory cycloreversion process. However, the barrier to the conrotatory cycloreversion in the photoexcited state remains low, meaning the reverse reaction can be triggered on-demand by irradiation in the absorbance window of the closed species. Indeed, another ideal feature of the DTE system is that the open forms absorb mainly in the UV, whilst the closed form possesses a distinct visible absorbance band. Photostationary states consisting of almost entirely open or closed isomers can therefore be achieved. Finally, to prevent competitive olefin isomerisation, the central double bond is most often constrained into a cyclic system to lock the molecule into the *Z* configuration and prevent conversion to the *E* form upon photoirradation.[217] These design features render DTE derivatives almost perfect molecular photoswitches and significant effort has been devoted to tune their behaviour towards a huge number of applications.[217–220]

5.2.2 Dithienylethenes as DNA-targeting molecules

Unlike the stiff-stilbene scaffold developed in the previous chapters, the possibility of deploying DTE as a scaffold for photoresponsive DNA-binding molecules has been explored briefly by others (Figure 5.2). However, to date, all these studies have focused exclusively on duplex DNA and the possibility of using DTE-based ligands to target G-quadruplexes has not yet been considered, despite the particular opportunities afforded by these particular nucleic acid structures.

In 2011, Feringa and co-workers reported the binding properties of the open and closed forms of DTE derivative **72** to self-complementary DNA hairpins.[221] The DTE core was derivatised with amine-terminated side chains which, when protonated at physiological pH, form electrostatic interactions with the DNA phosphate backbone. Both isomers were shown to bind duplex DNA, evidenced by induced circular dichroism in the absorbance bands of the open and closed ligand forms. These spectral features were used to monitor the reversible photoswitching between the ligand states in the presence of DNA upon alternate irradiation with UV and visible light. The binding of both ligand forms was shown to depend strongly on pH, with dissociation of the DNA/ligand complex observed at alkaline pH indicating poor association of the deprotonated ligand species for the DNA. This led the authors to propose that the ligands interact with DNA primarily through electrostatic binding modes to the DNA phosphate backbone, rather than intercalation with the base pairs. Though the chiroptical properties of the system were reversibly switched, this spectroscopic output only serves to demonstrate that the DTE can be isomerised in the presence of DNA. Evidence of

photocontrol of ligand activity (e.g. a change in binding mode and/or binding affinity) necessary for functional applications was not observed.

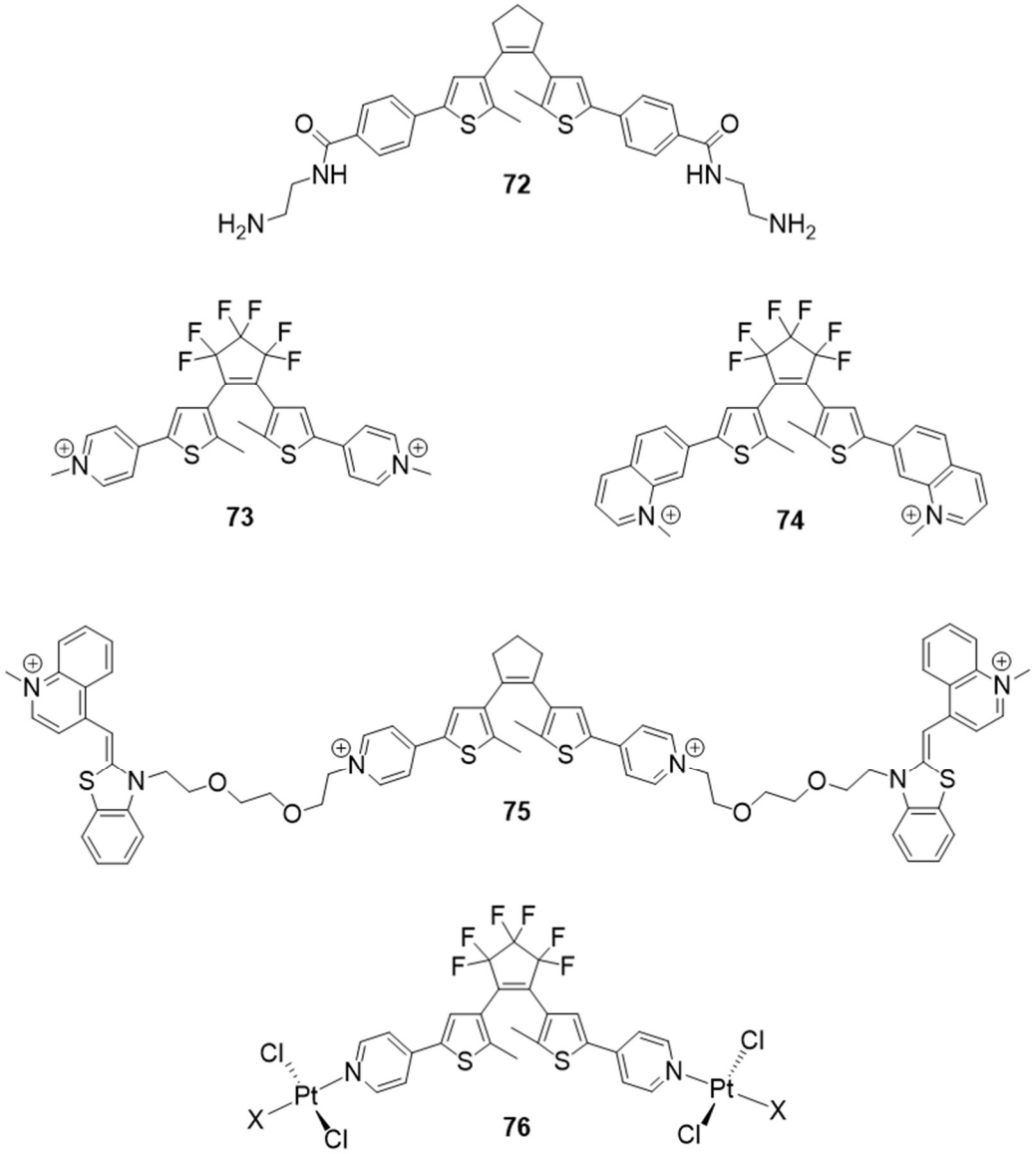


Figure 5.2: Examples of DTE switches investigated for binding to duplex DNA.

Andreasson and colleagues investigated the binding of DTE derivatives functionalised with *N*-methylated aromatic moieties (compounds **73** and **74**) to calf-thymus DNA.[222] Binding constants in the order of 10^5 M^{-1} were obtained in 5 mM NaCl solution. That the affinity was found to be weaker (though not explicitly quantified) in the presence of higher salt concentrations indicated electrostatic binding modes which are supressed in media of higher dielectric strength. Indeed, the quinolinium derivative **74** displays similar binding affinity to the

pyridinium analogue **73** suggesting that electrostatic interactions, rather than ligand size and shape, is the key determinant of binding affinity. A change in ligand binding mode upon photoisomerisation of **73o** to **73c** was inferred from linear dichroism experiments: the open form displayed a signature characteristic of intercalation, whilst the closed form appeared bind in the DNA groove.

Another interesting example came from the Yi group who exploited both the self-association and DNA binding properties of the thiazole orange dye to develop a DNA-gated photochromic DTE derivative, **75**. In the absence of DNA, the sandwiching of the thiazole moieties locks the molecule in the parallel conformation, both quenching the fluorescence and preventing the photocyclisation. Binding to DNA requires the molecule to adopt the photo-reactive antiparallel conformation, activating fluorescence, which can be switched off on-demand by isomerisation to the non-fluorescent closed form. This device was used to image DNA inside cells, with higher fluorescence observed in DNA-harbouring regions (the nucleus) versus locations were less DNA is present (the cytoplasm). Some inferences of the binding mode of **75** to self-complementary DNA duplexes were made using UV and CD spectroscopy, but the possible differences in binding modes between the open and closed forms were not part of the scope of this study.

Finally, Gamez studied the DNA binding of the open/closed forms of the DTE platinum complex **76** (Figure 5.2).[223] Through a fluorescence intercalator displacement assay, binding constants of the open and closed forms were estimated, with the closed form appearing to be 5-fold more affine for calf-thymus DNA (through comparison of the Stern-Volmer quenching constants). The increased affinity was proposed to result from the increased planarity and more compact structure of the closed isomer which translated into approximately 2-fold increase in cytotoxicity lung carcinoma cells.

5.2.3 Dithienylethenes as potential photoresponsive G4 ligands

Whilst the potential of DTE derivatives to act as photoresponsive DNA binding molecules has been considered previously, to date only binding to duplex DNA has been investigated. Surprisingly, no attempts to investigate DTE binding to G4 DNA structures have yet been made, despite the differences in secondary structure suggesting possibilities of additional binding modes where the state of the DTE photoswitch may exert more influence on binding. For example, the more rigid planar structure of the closed DTE isomers could be expected to stack more efficiently with G-tetrads than the conformationally flexible open isomers. As reported in Chapter 3, it was found that *N*-pyridinium stiff-stilbene ligands were particularly effective G4 binders, while discriminating effectively against duplex DNA, but the one-way photoresponse of the ligand ultimately limited the potential applications. In light of these

results, it was hypothesised that *N*-methylpyridinium DTE ligand **77** may combine the high G4 affinity demonstrated by the analogous stiff-stilbene compounds with more favourable photoswitching properties (Figure 5.3a).

Moreover, the possible influence of the cationic moieties appended to the DTE core in influencing G4 binding was also considered. Larger heterocyclic systems (e.g. quinolinium moieties) have been shown to be beneficial for G4 binding efficiency since they increase the surface area of the ligand, which is favourable for forming selective stacking interactions with G-tetrads. In particular, PhenDC3 (**10**) is a particularly effective and selective G4 ligand. Analogues of DTE **77** bearing quinolinium side-chains, **78** and **79**, were therefore desired to test examine the effects of the *N*-heterocycle on ligand activity (Figure 5.3b).

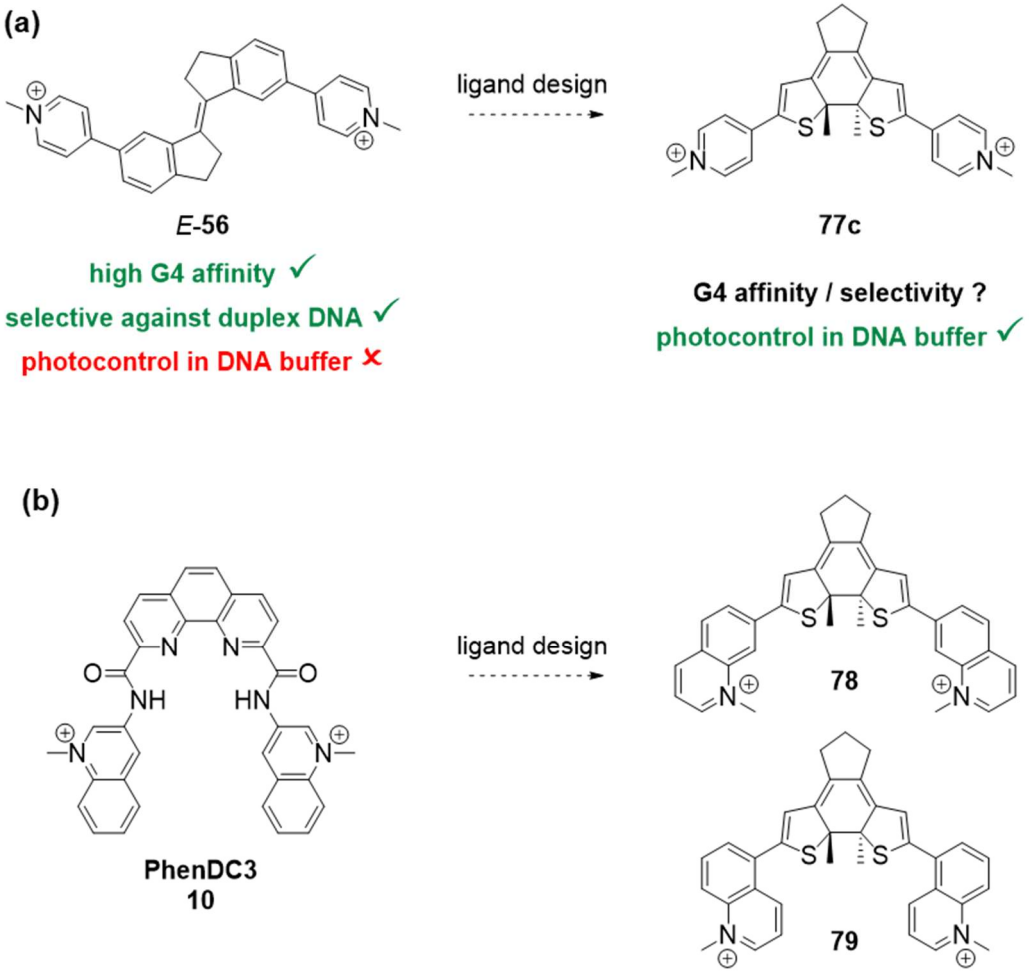
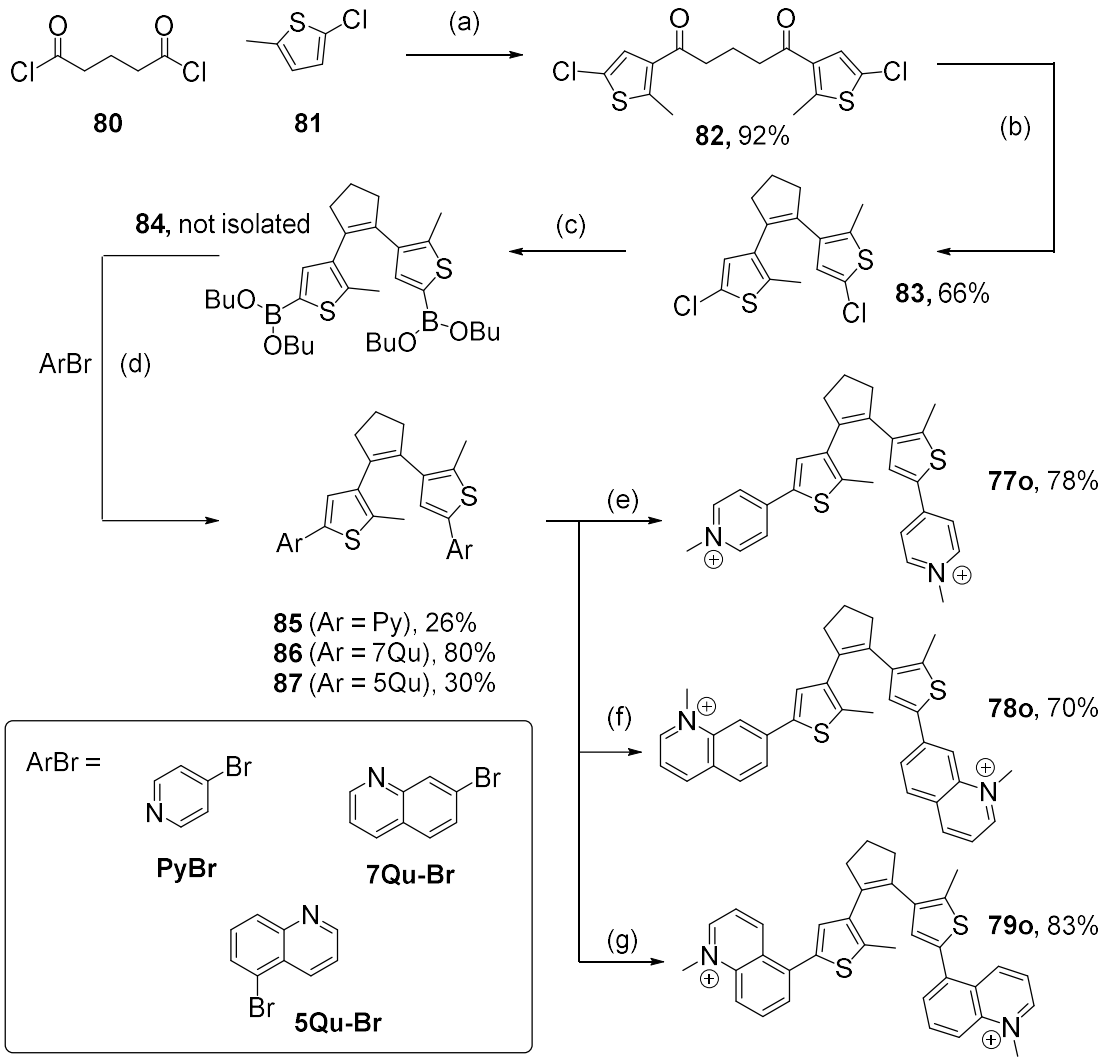


Figure 5.3: Design hypotheses explored in this chapter. (a) Rationale for investigation of *N*-methylpyridinium DTE as a G4 binding ligand based on previous results; (b) extension of *N*-methylated heterocycles to more closely mimic the structure of a known selective G4-binding chemotype.

5.3 Compound synthesis

5.3.1 Synthesis of the open ligand forms

The DTE derivatives were synthesised using established procedures to access key intermediate **83** (Scheme 5.1).[224] The synthesis of compound **77o** has been reported previously towards the development of imaging agents,[224] whist derivatives **78o** and **79o** are novel compounds. Friedel-Crafts acylation of 2-chloro-5-methylthiophene **70** with glutaryl chloride **81** afforded intermediate dione **82** in quantitative yield. Intramolecular McMurry coupling, discussed previously in Chapter 2, afforded the cyclised intermediate **83**. The pendant aryl fragments were then attached by Suzuki coupling (mechanistic discussion provided in Chapter 3) to the appropriate aryl bromide after *in situ* generation of the boronic ester **84**. The low yields obtained in the case of **85** and **87** were a result of the difficulty of removing the triphenylphosphine oxide impurity remaining after the coupling reaction. Finally, the compounds were methylated with iodomethane to afford the target compounds **77o-79o** as bis-iodide salts. The quinoline derivatives **86** and **87** appeared less reactive than pyridine **85,** requiring elevated temperatures to affect methylation whilst the pyridine compound reacted rapidly at room temperature.



Scheme 5.1: Synthetic route to DTE G4 ligands **77o**-**79o**. Reagents and conditions: (a) $AlCl_3$, DCM, r.t. 18 h; (b) Zn, $TiCl_4$, THF, 65 °C, 4 h; (c) ^{n}BuLi, THF, 0 °C, 30 min, then $B(O^nBu)_3$, THF, r.t. 1 h; (d) ArBr, $Pd(PPh_3)_4$ (10 mol%), K_2CO_3, THF, 70 °C, 18 h; (e) MeI, DCM, r.t.; 3 h (f) MeI, MeCN, 70 °C, 16 h; (g) MeI, DCM, 50 °C, 16 h. Compounds **86** and **78o** synthesised by Dr Javier Ramos Soriano.

5.3.2 Photogeneration of the closed isomers

With the open forms of the ligands **77o-79o** in hand, samples of the closed ligand forms were generated by irradiation of the samples using a domestic 320-400nm UV lamp. Though ultimately this light source was not suitable for experiments involving DNA, it proved an expedient and easily accessible way to perform the photoreaction on sufficient scale to allow the conversion to be determined accurately by NMR spectroscopy. These samples of known concentration and composition could then be used as stock solutions for the DNA binding

assays in order to compare the G4 binding activity of both isomers without the complication of performing the photoreactions in the presence of DNA in the first instance.

Under these conditions, a 1 mM solution of the open form of pyridine derivative (**77o**) cyclised very readily on exposure to UV light (Figure 5.4). 45 seconds irradiation was sufficient to promote 90% conversion and, after 5 minutes, quantitative conversion to the closed form was achieved.

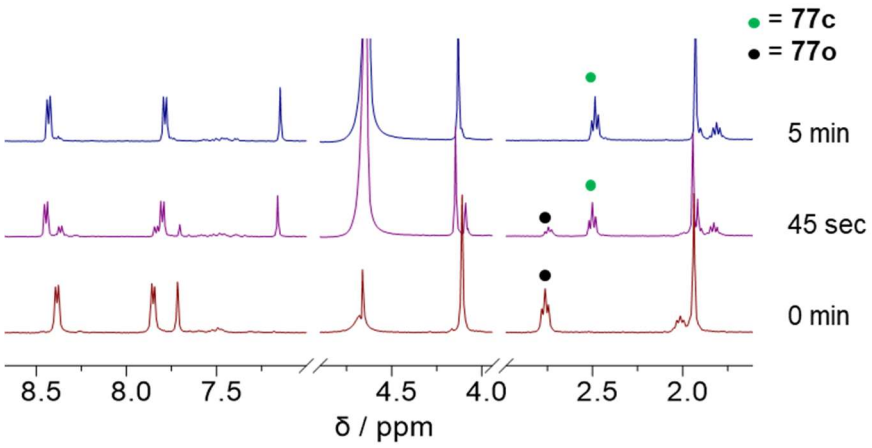


Figure 5.4. Cyclisation of **77o** to **77c** in D_2O upon exposure to UV (320 – 400 nm) light, monitored by ^1H NMR spectroscopy.

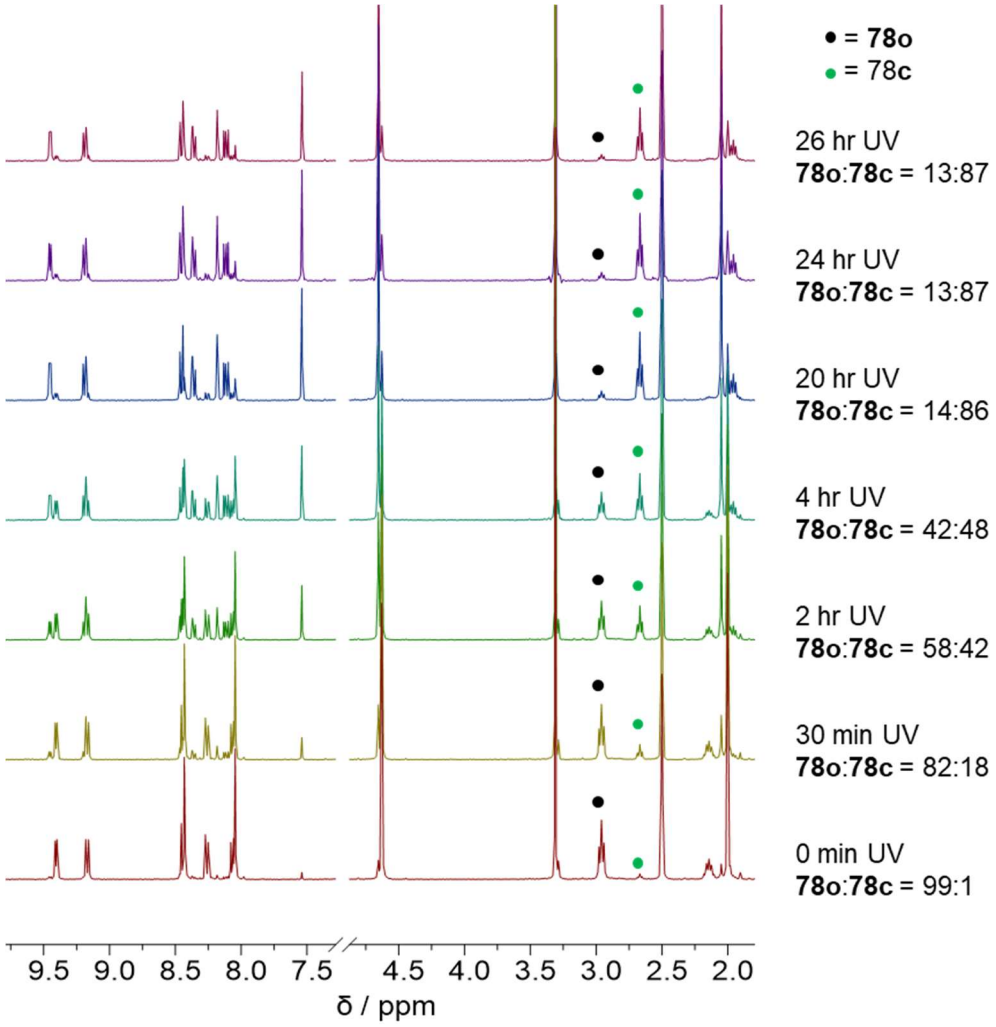


Figure 5.5. Cyclisation of **78o** to **78c** in DMSO upon exposure to UV (320 – 400 nm) light, monitored by ^{1}H NMR spectroscopy.

The novel quinolinium derivative **78o** was poorly soluble in water, so was instead isomerised in DMSO-d_6. Interestingly, the photocyclisation kinetics were much slower than those observed for the pyridinium analogue **77o** (Figure 5.5). Indeed, after 30 minutes irradiation only 18% conversion was detected and the sample had to be irradiated overnight to reach the photostationary state, corresponding to approximately 90% conversion to the closed form. Surprisingly, derivative, **79o**, did not undergo any cyclisation under the same conditions (Figure 5.6).

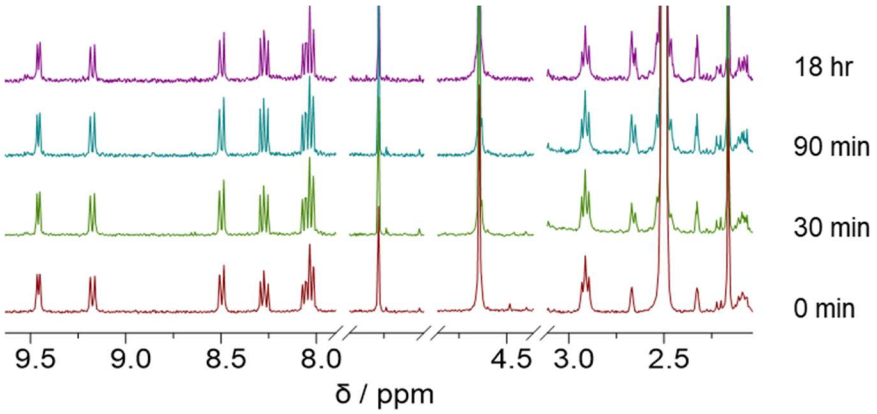


Figure 5.6: Attempted cyclisation of **79o** in DMSO upon exposure to UV (320 – 400 nm) light monitored by ^{1}H NMR spectroscopy. This demonstrates photostability of the open form even after long irradiation times.

5.4 Photoswitching studies

With reference samples of known composition of **77o/77c** and **78o/78c** (determined by NMR) in hand, efforts could turn to examining the photoswitching behaviour of the compounds in conditions relevant to DNA-based applications, namely 100 mM phosphate buffer at pH 7.4. First the photochromic behaviour of the compounds was investigated in the absence of DNA in order to develop suitable irradiation protocols compatible with DNA-based studies, before the effect of the DNA on the switching process was examined.

5.4.1 Reversible visible-light photoswitching in buffered aqueous solution

Since the broad-spectrum UV lamp used to perform the preparative photocyclisation of **77o** and **78o** was ultimately incompatible with irradiation of DNA samples and biological applications, photoirradiation experiments in buffered conditions were conducted using monochromatic visible light using 4.5 mW laser diode modules. Samples were irradiated under identical conditions (50 μM ligand in 100 mM sodium phosphate buffer, pH 7.4, 2 mL sample volume in a 3 mL cuvette, path length 10 mm) so that the systems could be compared under conditions relevant to the DNA binding assays.

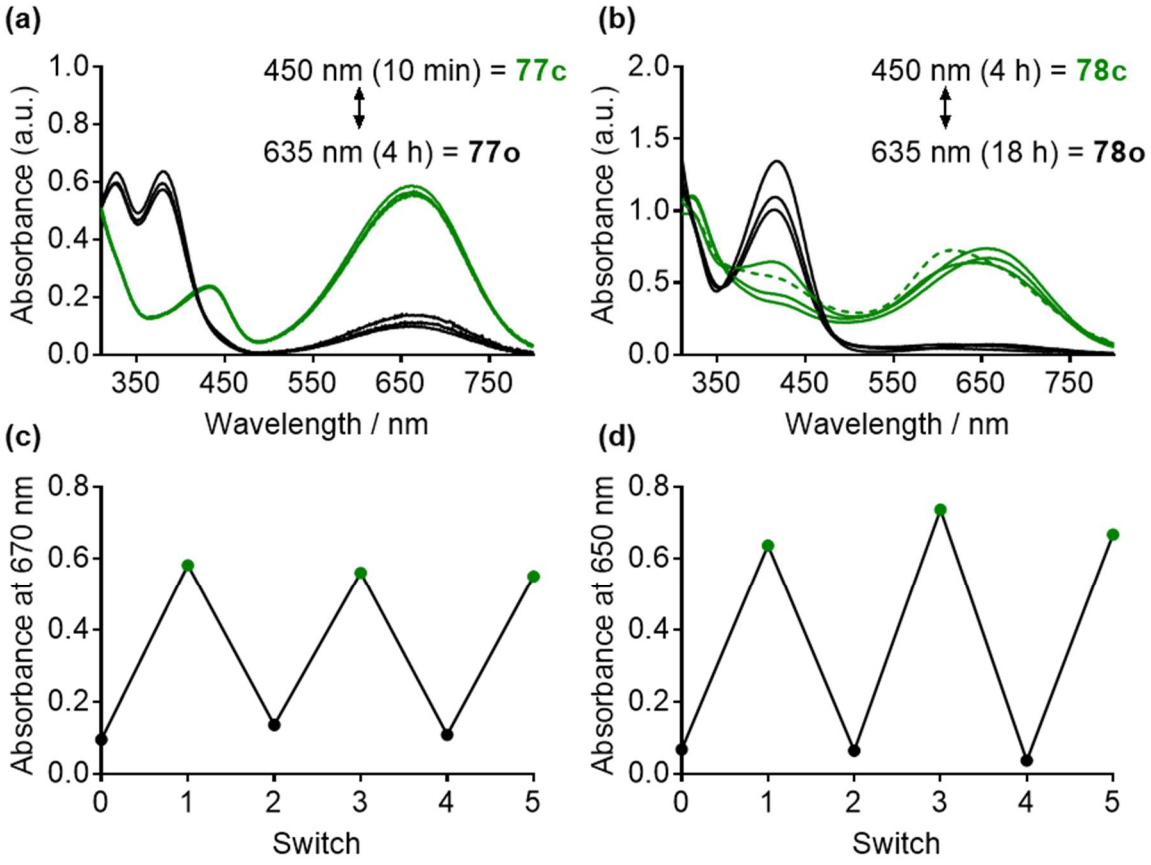


Figure 5.7. Reversible photoswitching of (a,c) **77o/77c** and (b,d) **78o/78c** in sodium phosphate buffer. (a-b) superimposed UV visible spectra of the samples obtained after alternate irradiation with 450 nm and 635 nm light; (b) photoswitching between open and closed forms monitored by red-shifted absorbance band of the closed isomers over five alternate switches. Ligand concentration 50 μM in 100 mM sodium phosphate buffer.

These concentrations were too low for NMR analysis to be practical, so the systems were monitored by UV/visible absorbance spectroscopy (Figure 5.7). Photocyclisation of both **77o** and **78o** was found to be promoted by 450 nm blue light, whilst 635 nm red light prompted the respective cycloreversion (Figure 5.7). The degree of conversion between the open and closed forms was estimated from the intensity of the red-shifted absorbance band of the closed isomer ($\lambda \approx 650$ nm). Since the open isomers do not absorb in this region, the intensity of this absorbance band could be normalised against spectra of samples of known composition in order to estimate the proportion of the closed isomer in the mixture, allowing the kinetics of the two photoswitches to be compared (Figure 5.8). In agreement with the NMR experiments in DMSO, the pyridinium-derived compound **77o** displays significantly faster photocyclisation kinetics, with full cyclisation prompted by 450 nm light in less than ten minutes (Figure 5.8a,

blue trace). Meanwhile, the quinolinium isomer **78o** required approximately 4 hours irradiation in order to generate a spectrum resembling the photostationary state in DMSO, (Figure 5.8a, red trace). In this case however, quantification of the extent of the ring closure reaction was complicated by the aggregation of the closed isomer under the buffered conditions. This effect led to a distortion in the reference spectrum obtained by diluting the 10 mM DMSO stock solution to 50 μM in buffer (Figure 5.7b, dotted trace). This effect was attributed to self-association of the ligand in aqueous conditions at the concentration range studied and was also observed in circular dichroism experiments (Section 5.5.2).

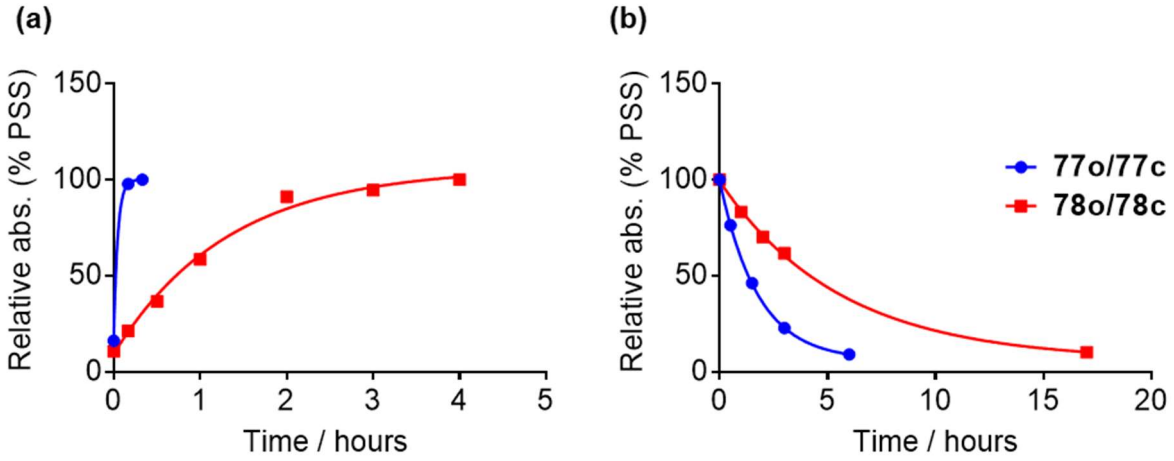


Figure 5.8. Kinetics of the (a) photocyclisation of **77o** and **78o** and (b) photocycloreversion of **77c** and **78c** in sodium phosphate buffer. The conversion was approximated by monitoring the formation/disappearance of the closed isomer and normalising the absorbance against a reference sample of the appropriate closed isomer. Ligand concentration 50 μM in 100 mM sodium phosphate buffer.

For both ligand pairs, the cycloreversion prompted by red light was slower than the corresponding ring closure (Figure 5.8b). Samples of pyridinium **77c** required approximately six hours irradiation to achieve near complete cycloreversion, though approximately 75% ring opening was observed after 3 h irradiation (Figure 5.8b, blue trace). Meanwhile, the quinolinium isomer **78c** demonstrated only 30% conversion after 3 h and required overnight irradiation to achieve complete ring opening (Figure 5.8b, red trace). Though these response times are rather slow, it should be noted that only low power (4.5 mW) irradiation sources were employed in these studies, and faster response times could easily be achieved by obtaining brighter light sources.

Finally, the reversibility of the switching process was examined by subjecting the ligands to alternate irradiation with 450 nm and 635 nm light (Figures 5.7c and 5.7d). Pleasingly, robust

and reversible photoswitching of the pyridinium analogue **77o/77c** was achieved. The spectra of the respective isomers remain virtually superimposable over irradiation cycles and a sharp isosbestic point is preserved indicating clean photoisomerisation between the two ligand states without decomposition. Reversible photoswitching of the quinolinium derivative **78o/78c** was also achieved, but the switching did not appear as efficient as for pyridinium derivative **77o/77c**. It is possible that the prolonged irradiation periods necessary to promote the isomerisation allow the accumulation of photodecomposition products, but the analysis is also complicated by the spectral distortion from the aggregation of ligand **78c** under the experimental conditions. Nonetheless, reversible switching between the two ligand forms is still demonstrated, even if it is not as efficient as for the pyridinium couple **77o/77c**.

*5.4.2 Photocyclisation of **77o** the presence of G4 and duplex DNA*

Having demonstrated the photoswitching of ligands **77o/77c** and **78o/78c** in the absence of DNA, the effect of the presence of different DNA secondary structures on the photocyclisation of **77o** (Figure 5.9) was examined. DNA was added to the samples of **77o** and the kinetics of the cyclisation were compared. In separate UV/visible titration studies (Section 5.5.3) it was observed that at 10 μM ligand concentration a G4 concentration of 35 μM was sufficient to result in full ligand binding to G4; the same DNA and ligand concentrations were employed here. The cyclisation of **77o** was too rapid under irradiation with 450 nm light for kinetic studies with the experimental setup available, so the samples in this case were irradiated under ambient light. Interestingly, the presence of both G-quadruplex structures examined (telo23-K^+ hybrid G4 and telo22-Na^+ antiparallel G4) markedly accelerated the cyclisation of **77o** relative to the kinetics observed in the absence of DNA. Indeed, the extracted rate constant in the presence of antiparallel telo22 G4-Na^+ (k_{isom} = 7.6 x 10^{-4} s^{-1}) was over five-fold higher than that observed in the absence of DNA (k_{isom} = 1.7 x 10^{-4} s^{-1}). Meanwhile, in the presence of duplex DNA (ds26), a less that 2-fold increase in the k_{isom} was observed, suggesting that the DNA-accelerated photocyclisation is specifically accelerated by binding to the four-stranded structures. These interesting results are worthy of further exploration though were not taken further in the present project due to time constraints.

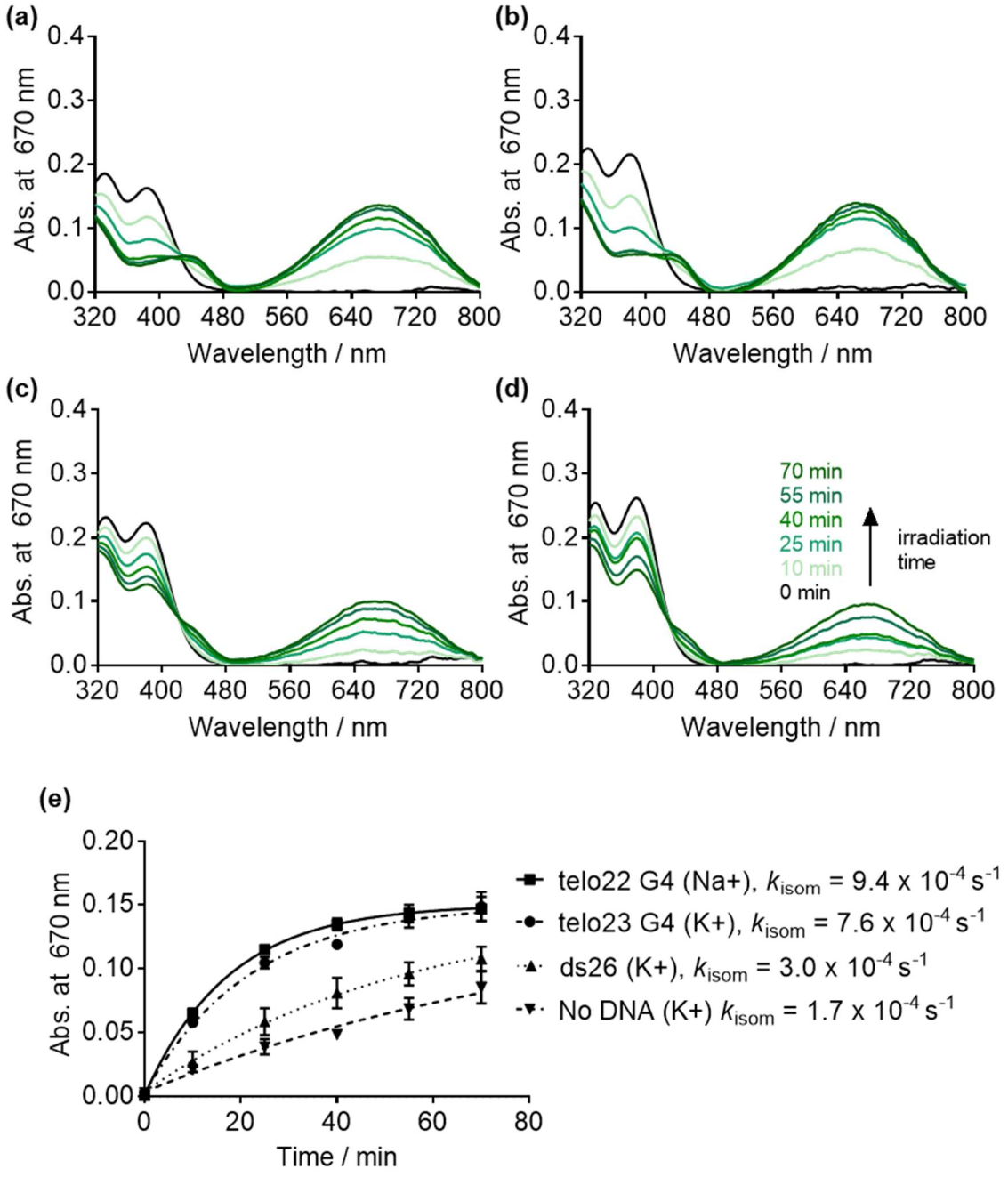


Figure 5.9. Effect of different DNA structures on the photocyclisation kinetics of **77o** to **77c.** Ligand concentration: 10 μM, DNA concentration: 35 μM in 100 mM sodium phosphate buffer. Samples were exposed to ambient light under identical conditions. (a-d) UV/visible spectra following irradiation for the times shown: (a) telo23-K^+ hybrid G4, (b) telo22-Na^+ antiparallel G4, (c) ds26 duplex DNA, (d) no DNA; (e) fitting of kinetic plots to first order rate equation to determine k_{isom} under the reaction conditions. The error bars show the deviations of two independent experiments.

5.5 Biophysical assays

With samples of the open and closed forms of ligands **77o/77c** and **78o/78c** in hand and confident that the ligands could be photoisomerised in conditions relevant to DNA folding, efforts turned to investigating the activities of the different compounds using the biophysical assays introduced in previous chapters.

5.5.1 FRET thermal melting assays

The ligands were initially screened for binding activity in the FRET thermal melting assay to provide insights into their ability to stabilise G4 DNA structures and consider the selectivity between G4 and duplex DNA.

5.5.1.1 Assay considerations

Two potential issues regarding the deployment of the FRET melting assay to study the DTE ligands were identified. The first was the possibility of unwanted photocyclisation of the two ligand forms under the assay conditions. However, at the excitation wavelength (λ_{ex} = 492 nm) employed in the assay, this was not expected to be problematic, since neither **77o** or **78o** absorb at this wavelength (Figures 5.7a and 5.7b). Of more concern was the potential for thermal cycloreversion of the closed isomers. The stability of the closed from at elevated temperature was therefore examined by UV/visible spectroscopy (Figure 5.10). Quinolinium ligand **78o** was shown to be stable at temperatures up to 90 °C with negligible cycloreversion observed at this temperature even at prolonged irradiation times (Figure 5.10b) Meanwhile, ligand **77c** appeared stable at lower temperatures but measurable cycloreversion was detected as the temperature was increased in the manner of the FRET assay. Therefore, the FRET results for this compound should be interpreted with caution since some isomerisation back to the open form could be expected to occur on the timescale of the experiment. Further isothermal techniques (CD spectroscopy, UV/visible spectroscopy and NMR spectroscopy) that are not vulnerable to this complication were later employed to more rigorously examine binding differences between the open and closed forms of the ligands.

5.5.1.2 Assay results

Ligands **77o/77c** and **78o/78c** were screened against three G4 models: the telomeric F21T G4 model under potassium- and sodium- rich conditions, and the oncogene promoter model FmycT. The self-complementary duplex hairpin was included to assess the selectivity of the ligands between G4 and duplex DNA.

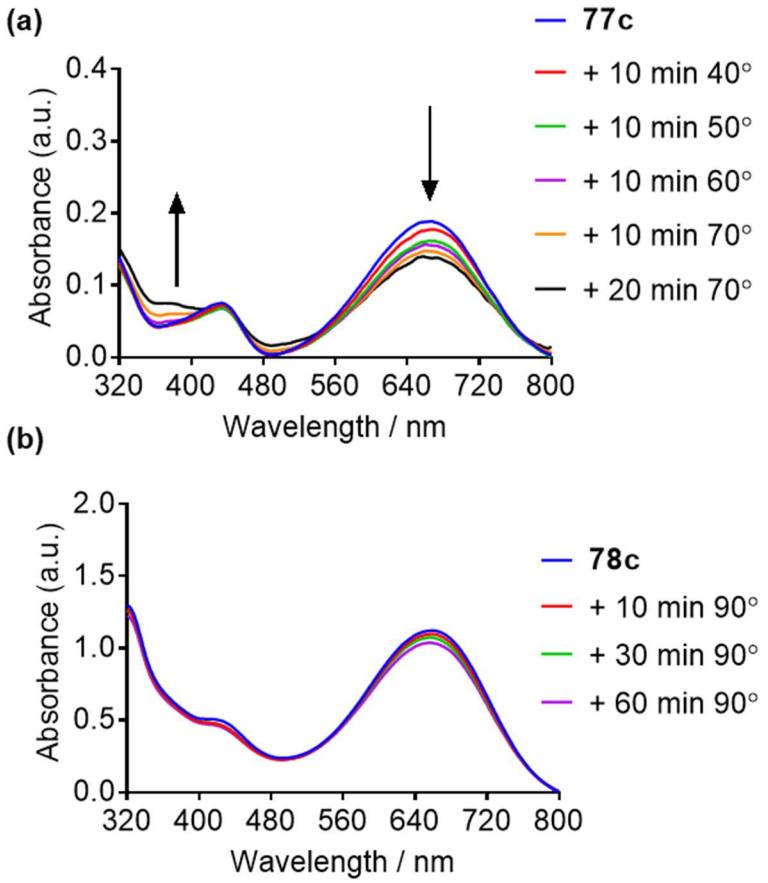


Figure 5.10: Investigations of the thermal stability of ligands (a) **77c** and (b) **78c** by UV/visible spectroscopy in water.

The results of the FRET assay are shown graphically in Figure 5.11 and representative raw data in Appendix A5 (Figures A5.1-A5.4). From the results, it can be seen that all ligands show selectivity for G4 DNA over the duplex model. Figures 5.11a-5.11c show measurable stabilisation of all G4 sequences is observed across the concentration range studied, whilst the stability of the duplex hairpin is comparatively unaffected even at high ligand concentrations (Figure 5.11d). For example, whilst ligands **77o**/**78c** induce stabilisation of all three G4 sequences at 2 μM (e.g. ΔT_m [F21T-K^+] = 15 °C for **78c**, Figure 5.11a), no stabilisation of the duplex DNA is observed at this concentration (Figure 5.11d). Indeed, even at 10 μM ligand concentration only slight (ΔT_m [F10T] = 4 °C) thermal stabilisation is induced in duplex DNA by **78c** (Figure 3.11d). This result suggests that DTE derivatives are better suited to binding G4 structures despite previous studies on the DNA binding of this scaffold being directed towards duplex DNA.

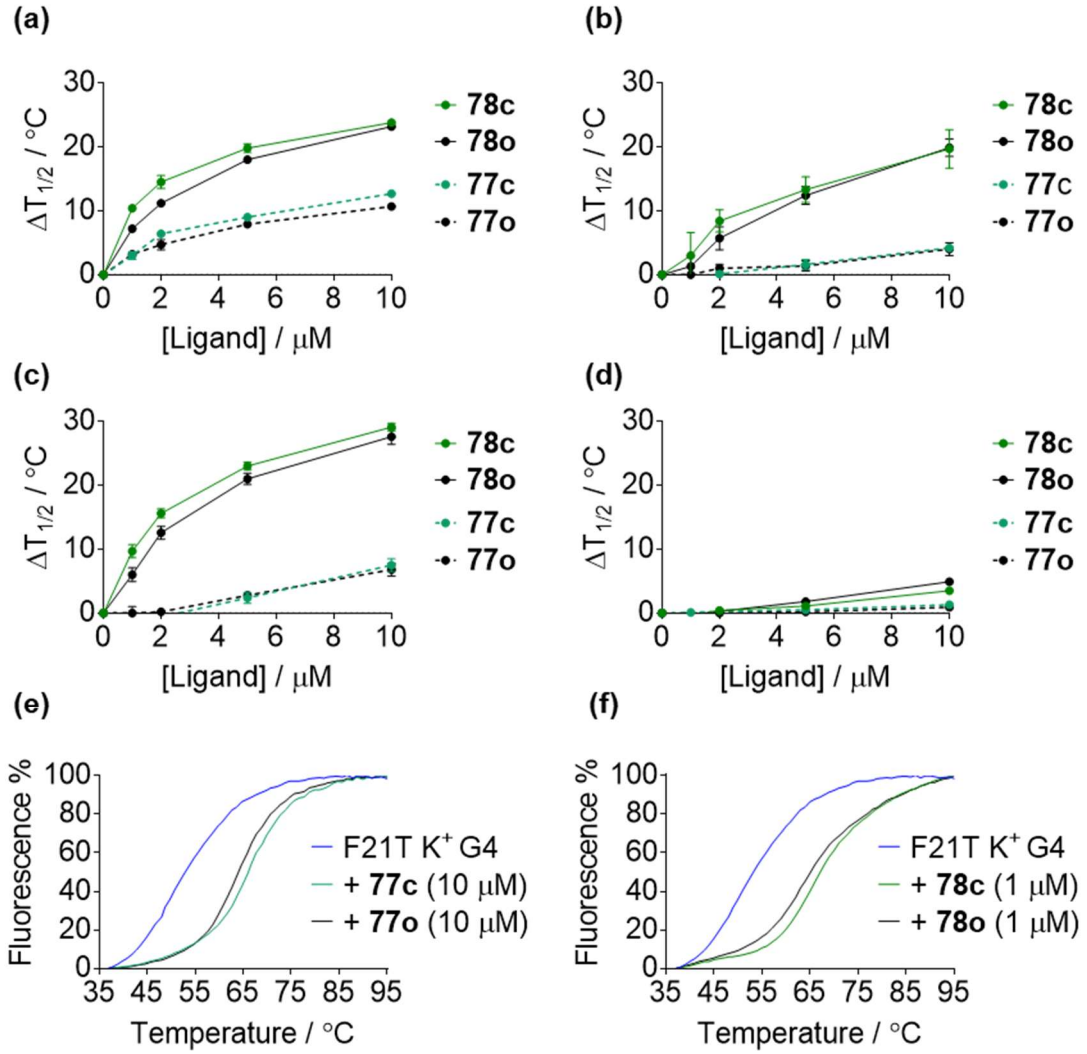


Figure 5.11: FRET thermal melting studies of ligands **77o/77c** and **78o/78c**. (a)-(d): thermal stabilisation values (ΔT_m) induced in (a) F21T-K^+ G4, (b) F21T-Na^+ G4, (c) FmycT G4, (d) F10T duplex. (e)-(f) comparison of representative melting curves of F21T-K^+ in the absence and presence of (e) **77o/77c** and (f) **78o/78c**

Another evident feature of the results is that the quinolinium ligands **78o**/**78c** induce significantly higher thermal stabilisation in the G4 sequences than the pyridinium analogues **77o/77c**. For example, the ΔT_m values induced in the F21T K^+ G4 by the quinolinium **78o/78c** ligands (ΔT_m = 8 - 10 °C at 1 μM ligand concentration) are approximately two- to three-fold higher than those recorded for the pyridinium ligands **77o/77c** (ΔT_m = 3 °C) under comparable conditions (Figure 5.11a). Even greater differences between the ligands are observed for the FmycT (Figure 5.11c) and F21T-Na^+ (Figure 5.11d) G4 models. These results suggest that G-

tetrad stacking interactions may be important in the binding of DTE ligands to G4 DNA structures since the more extended surface area of the quinoline appears to dramatically increase the ability of the ligand to stabilise the secondary structure.

Of perhaps more importance to the desired application of the DTE derivatives as photoresponsive G4 binding molecules is the fact that a difference in activity between the open and closed forms can be observed. This is particularly evident for the F21T model, in which both closed isomers induce higher thermal stabilisation in the G4 sequence than the corresponding open forms. For example, at 1 μM and 2 μM ligand concentration, the ΔT_m values are 3 – 4 °C higher for closed isomer **78c** than the open form **78o**. A similar effect is observed for the pyridinium ligands **77o**/**77c**, where the stabilisation values are in general 2 - 3 °C higher for the closed isomer. For both ligand pairs, the difference in activity can also be visualised by inspection of the raw melting curves which shows the greater thermal shift induced by the closed isomer at the same concentrations (Figures 5.11e and 5.11f). Though the differences in the ΔT_m between the isomers are rather modest, they were significant enough to be reproducibly measured and indicate that the closed isomer is the more active ligand form. The difference in activity of the ligand isomers was also observed for **78o**/**78c** against the FmycT sequence (Figure 5.11c) and the F21T-Na^+ sequence at lower concentration (Figure 5.11b). The pyridinium **77o**/**77c** isomers displayed similar activities against these G4 models, though the ΔT_m values are perhaps too low for the differences in activity to be evident in the case of this ligand pair.

The FRET results demonstrated DTE to be a viable scaffold for the design of new photochromic G4 ligands, demonstrating good ability to stabilise G4 with discrimination against duplex DNA. Furthermore, some difference in activity between the photochromic forms of both ligands was observed. Unfortunately, this activity difference was not sufficiently large to conclude these molecules could be deployed as photocontrolled G4 binders; possibly the change in ligand geometry upon photoisomerisation is insufficient to translate into large differences in activity. Nonetheless, in light of the possibility to photoisomerise the new DTE ligands without degradation in G4-relevant conditions, further studies into the binding modes of the photochromic ligand forms were therefore desired in order to obtain more direct evidence for the different ligand activities in light of these preliminary results.

5.5.2 Binding studies by CD and NMR spectroscopy

Though the FRET assays reveals some small differences in activity between the open and closed forms of ligands **77** and **78** the differences were not sufficiently large to draw strong conclusions, and more direct evidence of the differences in G4 binding between the respective isomers was desired. The potential of ligands **77o/77c** and **78o/78c** to influence the folding

dynamics of G4 structures was investigated by NMR and circular dichroism spectroscopy which together allow the systems to be explored in greater detail.

*5.5.2.1 Studies of N-methylquinolinium ligands **78o** and **78c***

Attempts were initially made to investigate quinolinium ligands **78o**/**78c**, since these derivatives appeared to the stronger G4 ligands based on the results of FRET assay. The telo22-Na^+ sodium structure was initially employed, since this sequence folds into a single well-characterised G4 conformation, allowing facile assignment of many of the imino and aromatic resonances from simple 1D ^1H spectra as discussed in previous chapters.

The closed isomer **78c** induces distortion of the CD spectrum baseline (Figure 5.12b) of the G4, which was attributed to aggregation of the ligand in this rigid planar form as also suggested by UV/visible studies (Figure 5.7b). A control titration of the ligand into buffer (with no DNA present) revealed the same distortion, indicating the effect is ligand-specific and not the result of induced CD upon binding to the chiral DNA (Appendix A5, Figure A.5.5). It is therefore difficult to draw inferences about the ligand binding, other than that the overall antiparallel G4 signature appears to be preserved. Meanwhile, the CD baseline appears stable upon titration with ligand **78o** and therefore the spectral differences can be reasonably attributed to ligand binding events (Figure 5.12a). In this case, the ligand induces hypochromicity in the positive band at 240 nm and the negative band at 260 nm. These effects suggest that the ligand induces a disruption of the folded topology, possibly arising from an intercalative binding mechanism. NMR experiments of telo22 titrated with ligand **78o** reveal chemical shift perturbations of the resonances corresponding to the imino protons of the top (G4, G16, G21) and middle G-tetrad (G9, G15 and G3) suggesting that the ligand is able to intercalate the DNA structure between these two tetrads whilst resonances corresponding to the lower tetrad (G2, G10, G22) are not perturbed (Figure 5.12c). As the titration progresses the G8 resonance disappears, possibly suggesting disruption of the top G-tetrad by the binding of ligand **78o**, which may also explain the observed hypochromic effects in the CD signature. Meanwhile, the aromatic resonances of telo22 broaden slightly upon ligand binding but otherwise remain well-resolved and are largely comparable to the resonances observed in the absence of ligand.

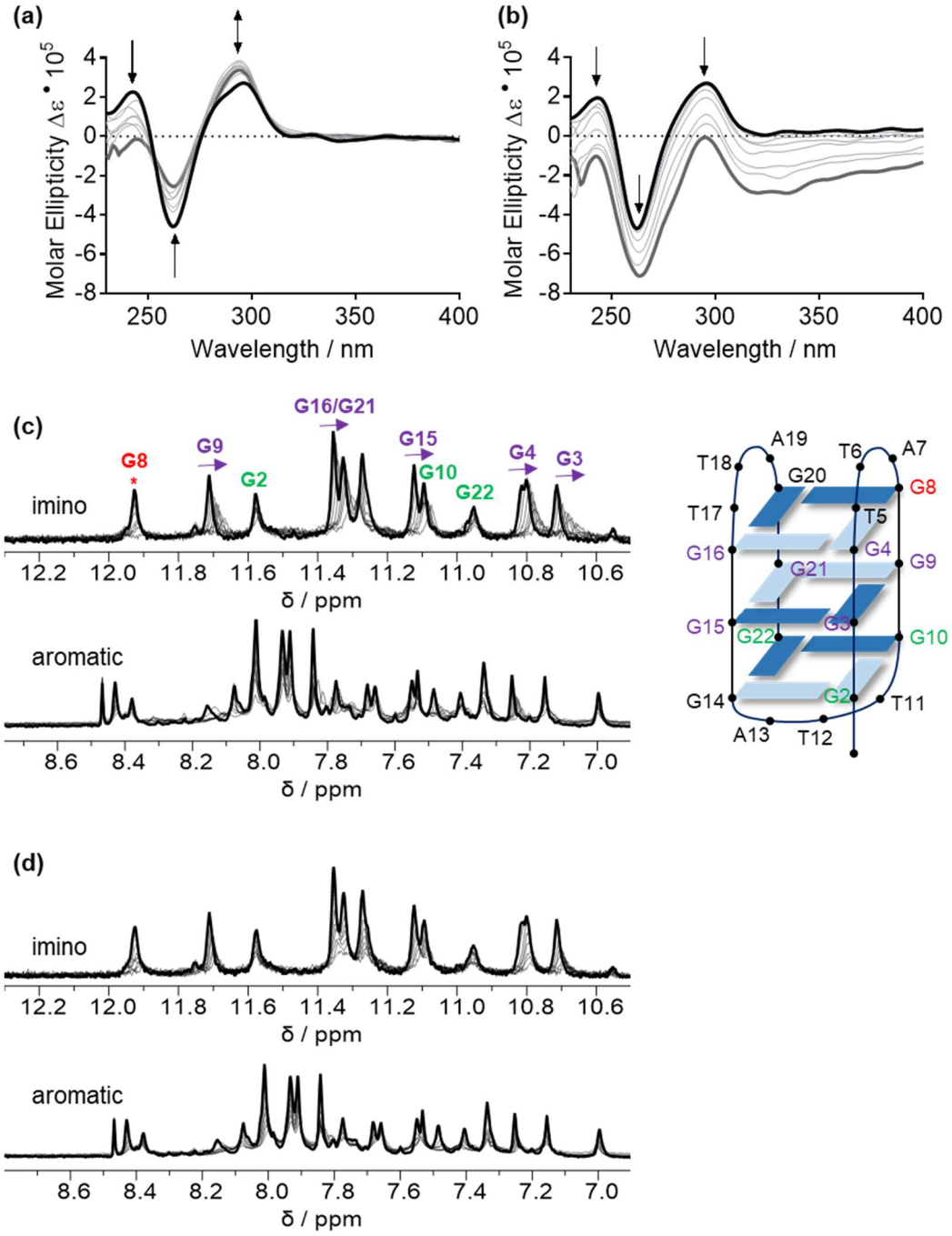


Figure 5.12: CD and NMR investigations of the binding of **78o/78c** to telo22-Na^+ G4 (a)-(b) CD spectra of telo22 in the absence (black trace) and presence (dark grey trace) of 7 equiv. ligand (a) **78o** and (b) **78c** (DNA concentration 4.2 μM), intermediate titration points are shown as light grey traces; (c)-(d) overlaid imino and aromatic NMR spectra of telo22 (black) and increasing concentrations (grey) of ligand (c) **78o** and (d) **78c** up to 2 equiv. ligand (DNA concentration 185 μM). The cartoon representation in (c) identifies the resonances that disappear (red), are perturbed (purple) and not perturbed (green) by ligand **78o**.

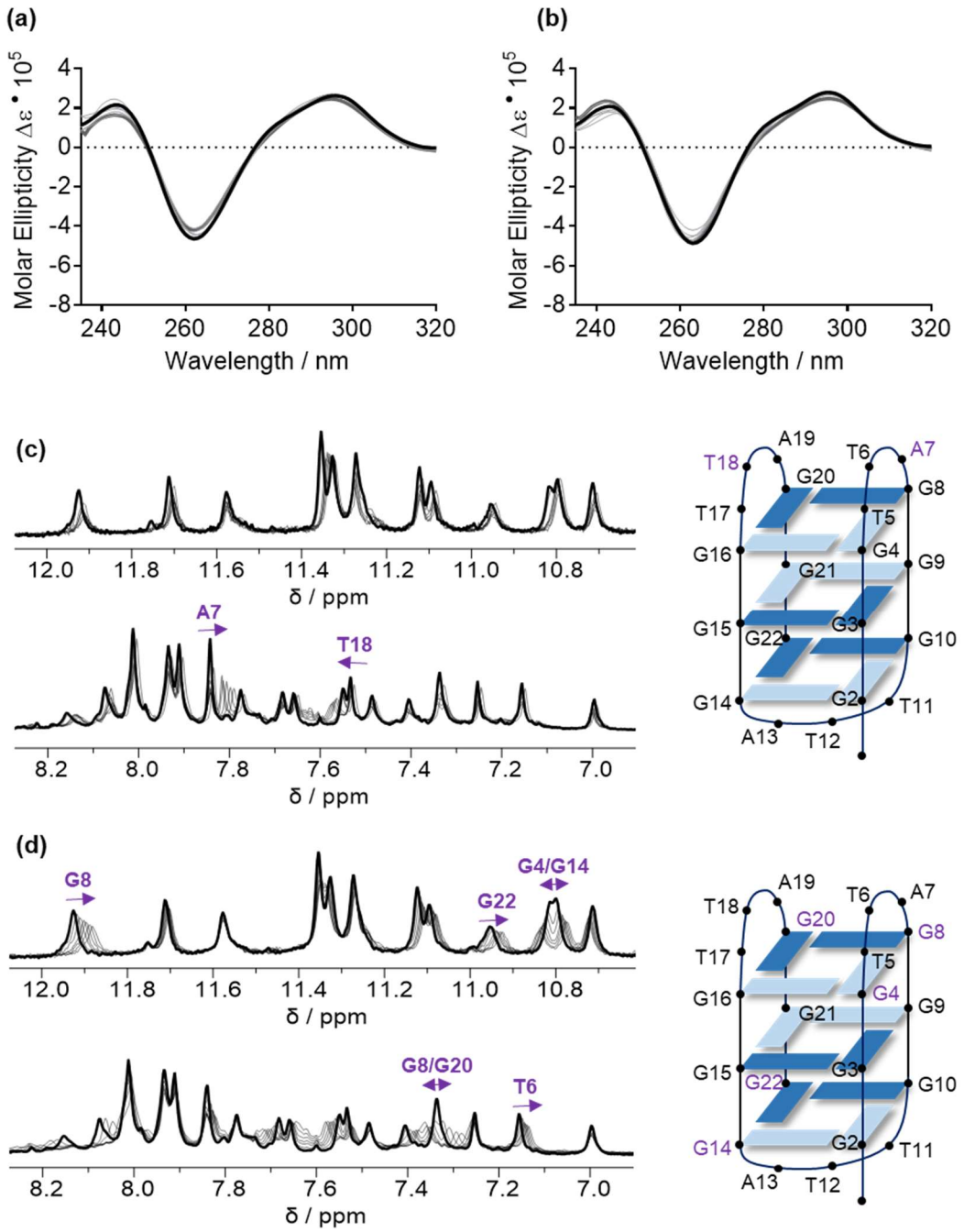


Figure 5.13 CD and NMR **i**nvestigations of the binding of **77o/77c** to telo22-Na^+ G4 (a)-(b) CD spectra of telo22 in the absence (black trace) and presence (dark grey trace) of 7 equiv. ligand (a) **77o** and (b) **77c** (DNA concentration 4.2 μM), intermediate titration points are shown as light grey traces; (c)-(d) overlaid imino and aromatic NMR spectra of telo22 (black) and increasing concentrations (grey) of ligand (c) **77o** and (d) **77c** up to 2.5 equiv. ligand (DNA concentration 185 μM). The cartoon representations in (c) and (d) identifies the resonances that are perturbed (purple) on addition of ligand.

In contrast to the titration with **78o**, the closed isomer **78c** does not cause any specific chemical shift perturbations of the telo22 imino resonances (Figure 5.12d) and all signals appear to broaden and reduce in intensity in a similar manner. Given that the CD studies indicate ligand aggregation even at comparatively low concentration, it is likely this effect complicates the behaviour of the ligand in the NMR studies resulting in the non-specific broadening observed. Therefore, the binding mode of this form of the ligand could unfortunately not be reliably compared with ligand **78o**.

*5.5.2.2 Studies of N-methylpyridinium ligands **77o** and **77c***

Neither form of ligand **77o**/**77c** induced perturbations in the CD spectrum of telo22 DNA (Figures 5.13a and 5.13b). This was not surprising, since the association of ligand with the telomeric G4 in sodium-rich conditions was indicated to be rather weak in the FRET studies. Nonetheless, weaker binding can still be observed in the NMR studies, as the concentrations employed are higher. Ligand **77o** does not significantly change the chemical shifts of the imino resonances (Figure 5.13c) but does induce specific perturbations of the aromatic resonances of bases A7 and T18 corresponding to the capping residues of the lateral loops above the top G-tetrad. That all residues of the lower G-tetrad are unperturbed suggests that ligand **77o** binds exclusively to the top of the G4 with 1:1 stoichiometry.

Meanwhile, the closed isomer **77c** causes clear perturbations to the chemical shifts of imino resonances (Figure 5.13d): G8 is shifted upfield whilst several aromatic proton resonances corresponding to the top G-tetrad are also perturbed, particularly G8 and G20. Significant shifts can also be observed for imino protons G22 and G14 corresponding to the lower G-tetrad, implying that ligand **77c** binds with a 2:1 stoichiometry to the upper and lower faces of the G4.

Pleasingly, the binding mode of the ligand could be reversibly switched several times by *in situ* photoisomerisation of the ligand. The sample was irradiated alternately with blue (450 nm) or red (635 nm) light and the switch in binding mode monitored by the chemical shift of the G8 and G4 imino resonances. The switching process was shown to be fully reversible with no observable degradation of the nucleic acid (Figure 5.14).

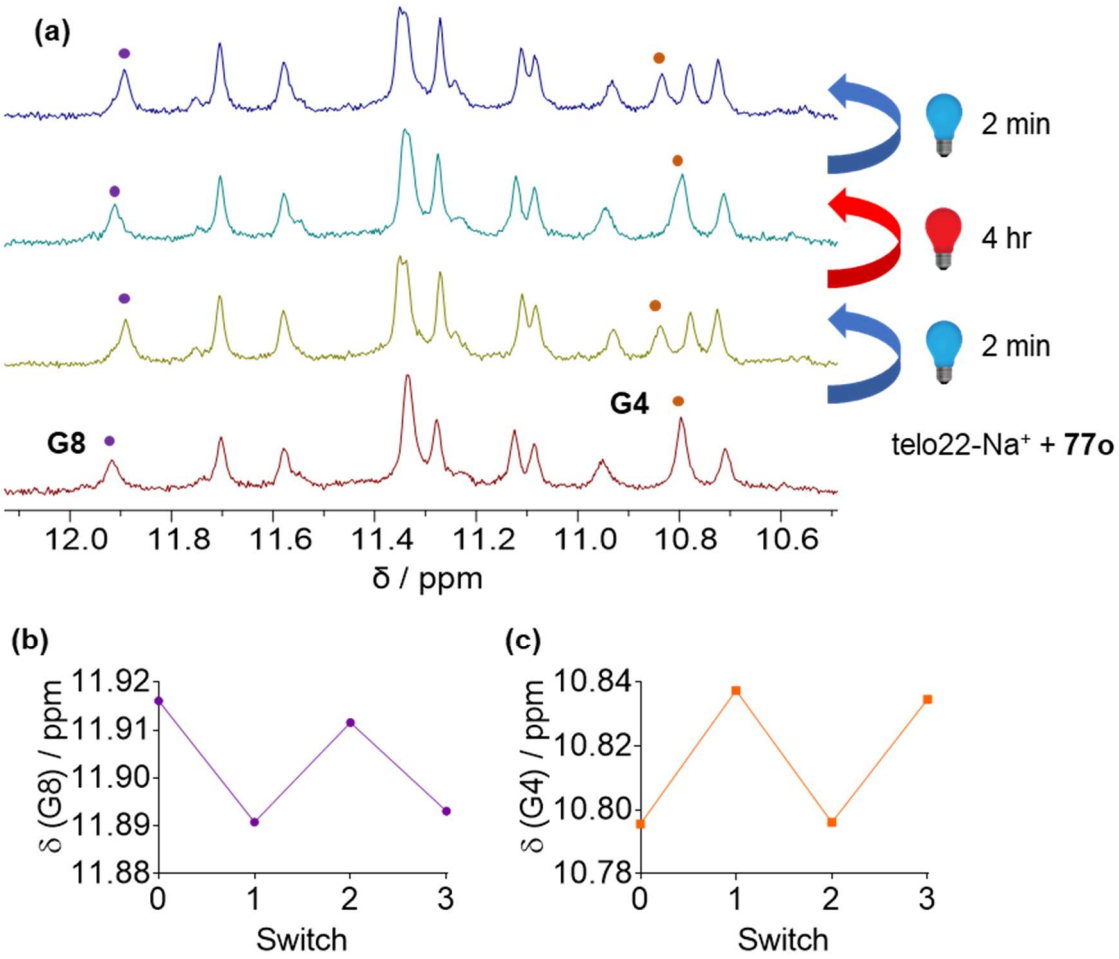


Figure 5.14: Reversible switching of binding mode of **77o/77c** to telo22-Na^+ G4. (a) NMR spectra of telo22 (DNA concentration 185 μM) and 1 equiv. ligand **77o** before and after each subsequent irradiation dose; (b)-(c) monitoring the change in ligand binding mode *via* the observed chemical shift of the (b) G8 and (c) G4 imino resonances.

Encouraged by these results, the effect of the **77o/77c** couple on the telo23-K^+ hybrid form of G4 DNA also investigated (Figure 5.15, stacked spectra in Appendix A5, Figure A5.6) since this sequence has greater physiological relevance owing to the higher concentration of potassium ions in cellular environments. As in the case of the antiparallel telo22-Na^+ G4, the closed isomer **77c** causes significant chemical shift perturbations (Figure 5.15a) of the imino signals of several guanine residues, indicative of stacking with the exposed G-tetrads. Again, since signals corresponding to both upper (G9/G21) and lower (G15/G23) tetrads are perturbed, a 2:1 ligand binding stoichiometry was inferred in this case. The intensity of the imino signals is maintained throughout the titration and the circular dichroism spectrum of the oligonucleotide is unaffected (Figure 5.15b) indicating that the binding of **77c** takes place

without disruption of the hybrid G4 fold. In sharp contrast, ligand **77o** induces the pronounced attenuation of many of the G4 imino signals, particularly G9, G10, G21 and G22 (Figure 5.15a, stacked spectra in Appendix A5, Figure A5.6). Indeed, only a small number of distinctive imino environments can be seen in the spectrum of the telo23/**77o** complex, rather than the 12 signals expected for a fully-folded G4 (if present as a single species). This suggests an intercalative binding mechanism for the open form of the ligand, requiring partial disruption of the G-tetrad network. Indeed, while the CD spectrum of telo23-K^+ is unperturbed by the addition of ligand **77c,** the addition of **77o** results in attenuation of ellipticity from 240-280 nm, also suggesting the disruption of the native G-tetrad network by addition of ligand (Figure 5.15c).

In a similar way to telo22-Na^+ these different effects of the **77o**/**77c** couple could be reversibly toggled by photoirradiation (Figure 5.16). Pleasingly, alternate irradiation with blue and red visible light allowed the switching process to be achieved at least 7 times. The degree of the conversion was followed by the integral of the G9/G10 imino resonances and was shown to depend on the irradiation time (Figure 5.16b). The regeneration of the CD intensity upon photocyclisation of **77o** to **77c** with blue light was also demonstrated (Figure 5.16c).

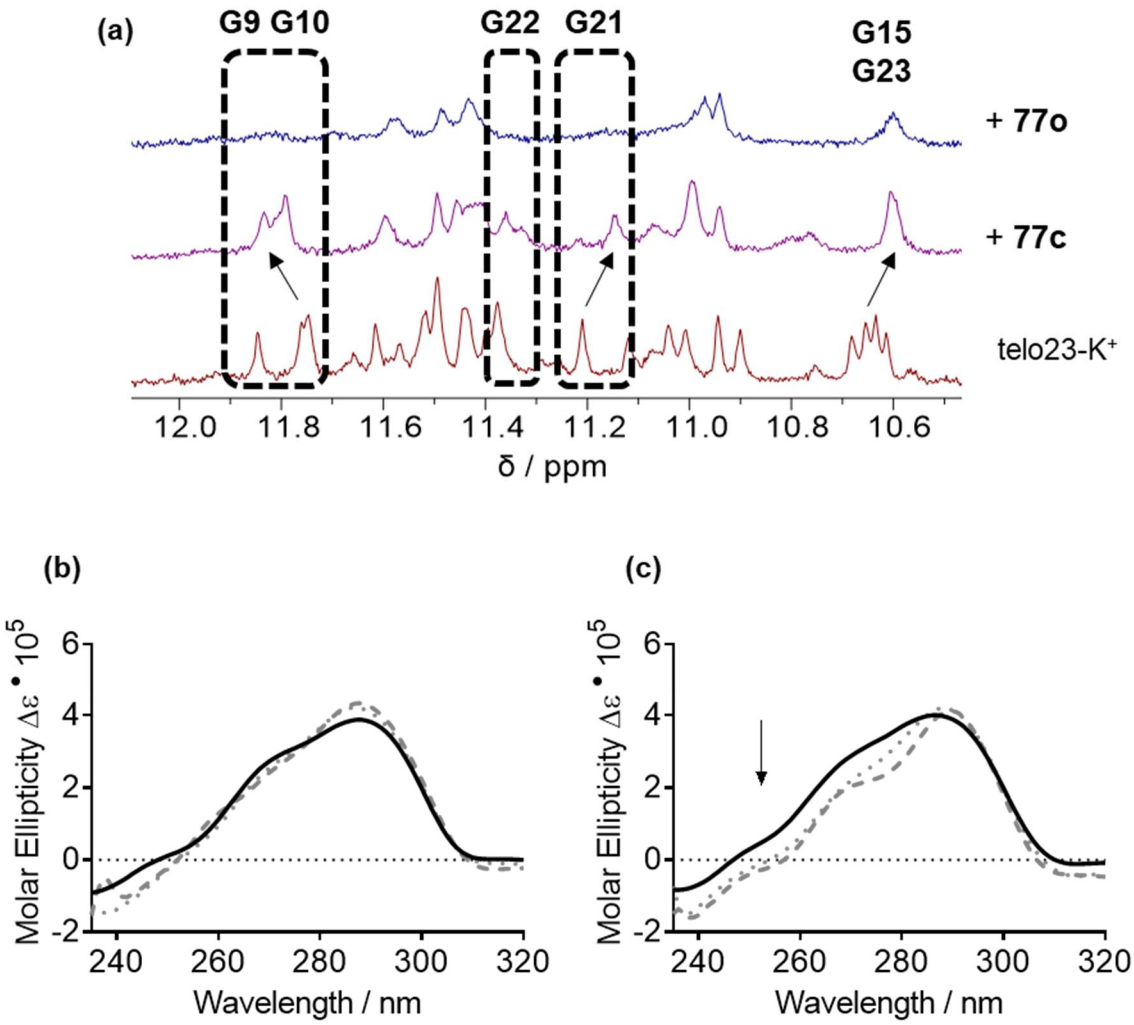


Figure 5.15. NMR and CD investigations of the binding of **77o/77c** to hybrid telo23-K^+ G4. (a) spectra of telo23 in absence of ligand and the presence of (middle) 2.5 equiv. **77c** and (top) 2.5 equiv. **77o**. DNA concentration = 185 μM; (b)-(c) CD spectra of telo23-K^+ in the absence (black trace) and presence of 50 μM (dotted grey trace) and 100 μM (dashed grey trace) (b) **77c** and (c) **77o** (DNA concentration 4.2 μM).

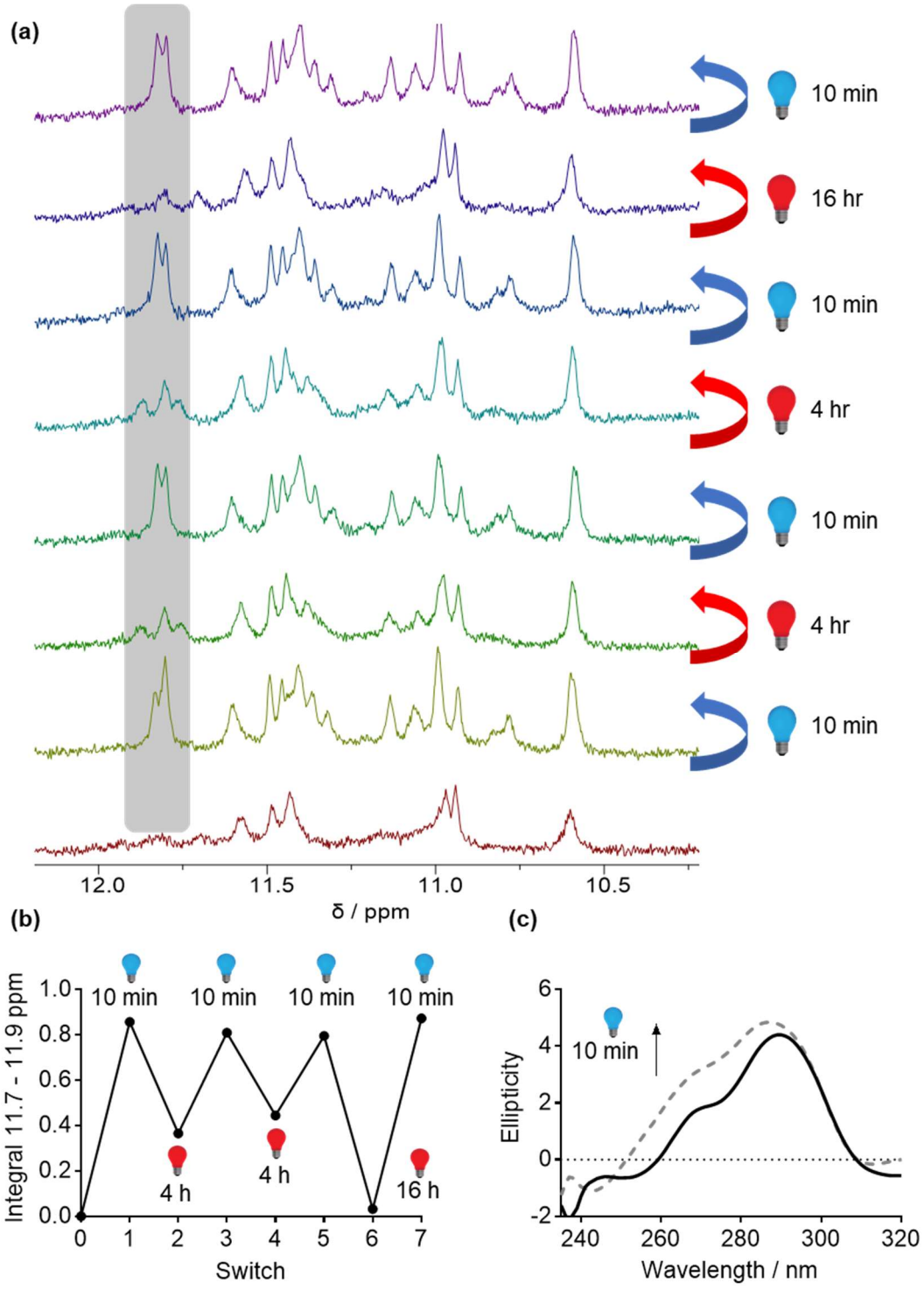


Figure 5.16: Reversible switching of binding mode of **77o/77c** to telo23-K^+ G4. (a) stacked NMR spectra of the DNA/ligand complex after each successive irradiation dose; (b) monitoring the change in ligand binding mode by the normalised integral of the shaded region in (a); (c) recovery of CD spectral intensity upon irradiation of the telo23/**77o** complex with blue light.

5.5.3 UV/visible titration absorbance titrations

Having demonstrated the ligand binding mode to G4 DNA could be controlled *in situ* for the ligand **77o/77c** couple, and in light of the small difference in activities observed between the open and closed ligand forms observed in the FRET assays, attempts were made to quantify the binding affinities of the ligands to antiparallel and hybrid telomeric DNA sequences using UV/visible titration studies in a manner to previous chapters.

5.5.3.1 Titrations with antiparallel (telo22-Na^+) G4

Considering first the *N*-methylpyridinium ligands **77o/77c**, binding of both open and closed forms to telo22 G4 is evidenced by hypochromicity in the ligand absorbance upon binding to DNA (Figure 5.17 and Table 5.1). For ligand **77c** this is accompanied by a bathochromic shift of approximately 9 nm, which may be attributed to the stacking with the upper and lower G-tetrads as observed in the NMR studies. The isotherms were fit to the independent and equivalent sites binding model (Eqn. (3.1), introduced in Section 3.3.2) assuming the stoichiometry suggested by the NMR studies, and similar binding constants (~ 1.5 x 10^5 M^{-1}) for both isomers were found. It is perhaps surprising that the different binding modes occur with the same apparent affinity given the different binding modes observed in the NMR titration studies, but this observation nonetheless appears to corroborate the results of the FRET study which indicates the two isomers induce a similar thermal stabilisation in the F21T-Na^+ G4 model (Figure 5.11b).

Turning attention to ligands **78o/78c**, the stronger binding of these derivatives to antiparallel telo22-Na^+ is observed (Figure 5.18), in agreement with the FRET results which indicate these induce greater thermal stabilisation of G4 at lower concentrations. The DNA-induced hypochromicity is larger than for the corresponding pyridine analogues, greater bathochromic shifts are observed and the fitted K_a values, (8.7 ± 2.9) x 10^5 M^{-1} for ligand **78o**, are higher (Table 5.1). Again, the open and closed ligand forms appear to display comparable affinity for G4, with ligand **78c** displaying an apparent K_a of (5.4 ± 0.04) x 10^5 M^{-1}.

5.5.3.2 Titrations with hybrid (telo23-K^+) G4

In the case of the telo23-K^+ G4 both ligands, stronger binding of all the ligands to the hybrid G4 was suggested by larger hypochromicities and bathochromic shifts in the spectra (Table 5.2). Despite this, the association constants calculated were similar to those observed for the telo22-Na^+ quadruplex and in both cases, with the exception of ligand **78c** which displays a K_a approximately two-fold higher, (11.2 ± 0.4) x 10^5 M^{-1} for the hybrid telo23-K^+ G4, than the antiparallel telo22-Na^+ G4 model. Again, the binding constants of the respective open and

closed ligand forms are similar, which is surprising in light of the different effects observed in the NMR titration studies for **77o/77c**.

Table 5.1: Summary of spectral changes observed for ligands **77o/77c** and **78o/78c** after addition of 3 equiv. antiparallel telo22-Na^+ G4 DNA and fitted K_a values using the stoichiometry indicated. The K_a values are the average of two independent experiments.

Ligand	Hypochromicity / %	Red shift / nm	Apparent K_a / 10^5 M^{-1}	Stoichiometry
77o	17	0	1.6 ± 0.2	1
77c	10	9	1.7 ± 0.8	2
78o	26	11	8.7 ± 2.9	3
78c	21	40	5.4 ± 0.04	3

Table 5.2: Summary of spectral changes observed for ligands **77o/77c** and **78o/78c** after addition of 3 equiv. telo23-K^+ hybrid G4 DNA and fitted K_a values using the stoichiometry indicated. The K_a values are the average of two independent experiments

Ligand	Hypochromicity / %	Red shift / nm	Apparent K_a / 10^5 M^{-1}	Stoichiometry
77o	35	5	1.5 ± 0.2	2
77c	21	12	2.1 ± 1.3	2
78o	35	15	8.1 ± 0.9	3
78c	15	56	11.2 ± 0.4	3

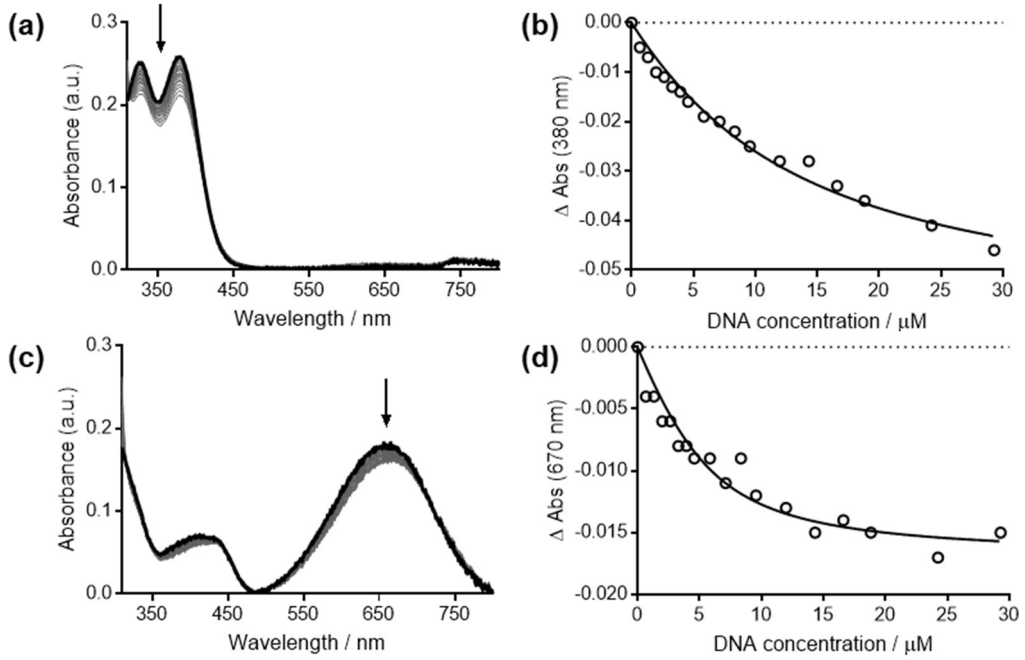


Figure 5.17: UV/visible titration studies of ligands **77o** and **77c** with antiparallel telo22-Na^+. (a),(c): titrations of 10 μM (a) **77o** and (c) **77c** with; (b),(d): extracted binding isotherms.

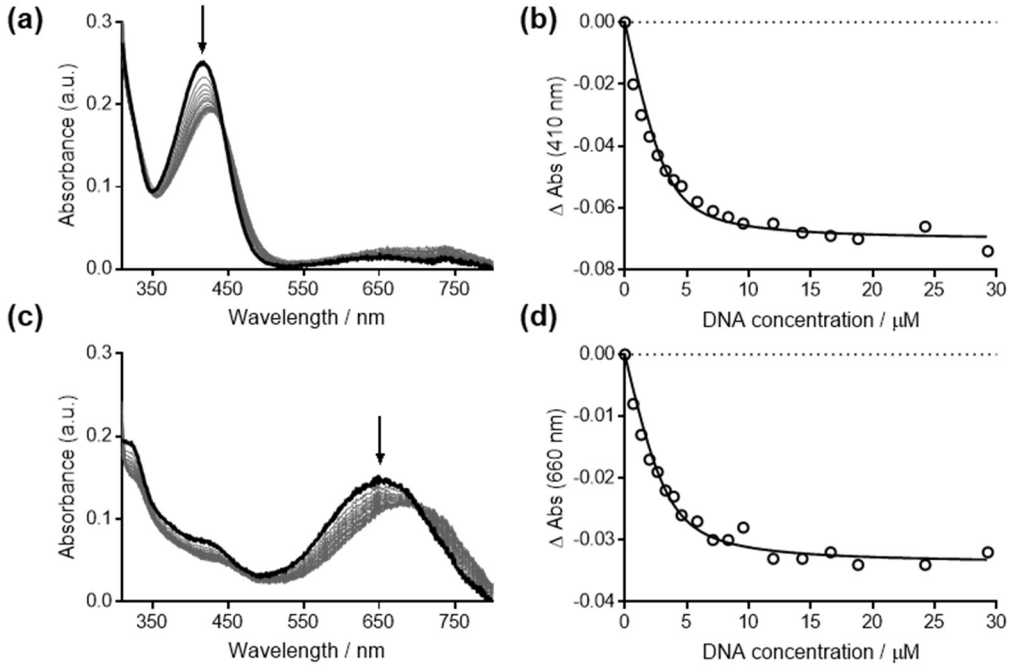


Figure 5.18: UV/visible titration studies of ligands **78o** and **78c** with antiparallel telo22-Na^+. (a),(c): titrations of 10 μM (a) **78o** and (c) **78c**; (b),(d): extracted binding isotherms.

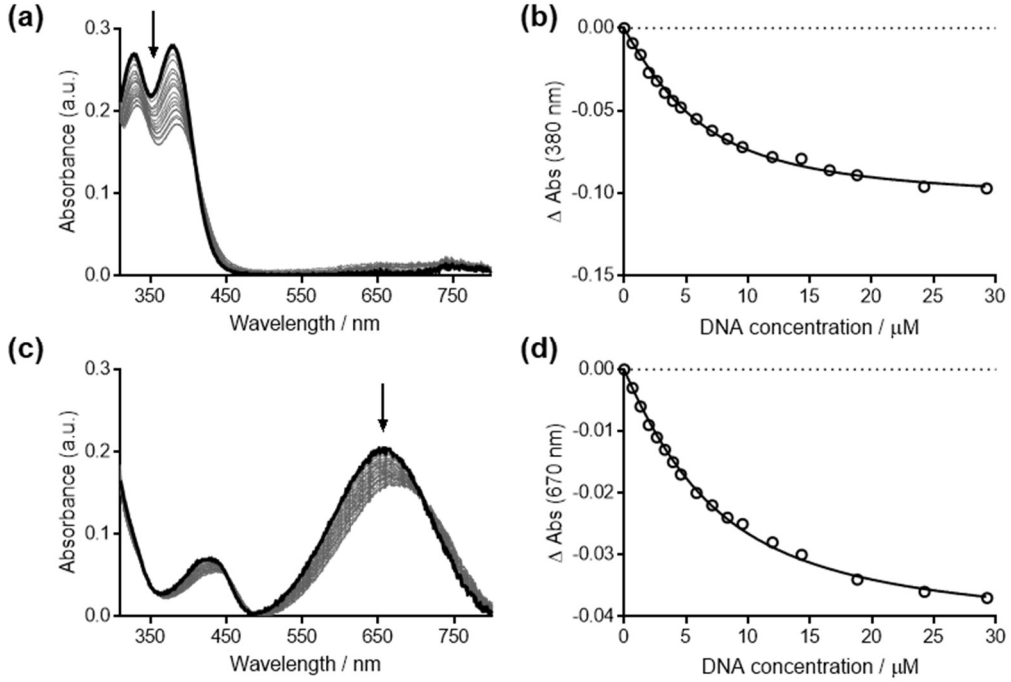


Figure 5.19: UV/visible titration studies of ligands **77o** and **77c** with hybrid telo23-K^+. (a),(c): titrations of 10 μM (a) **77o** and (c) **77c**; (b),(d): extracted binding isotherms.

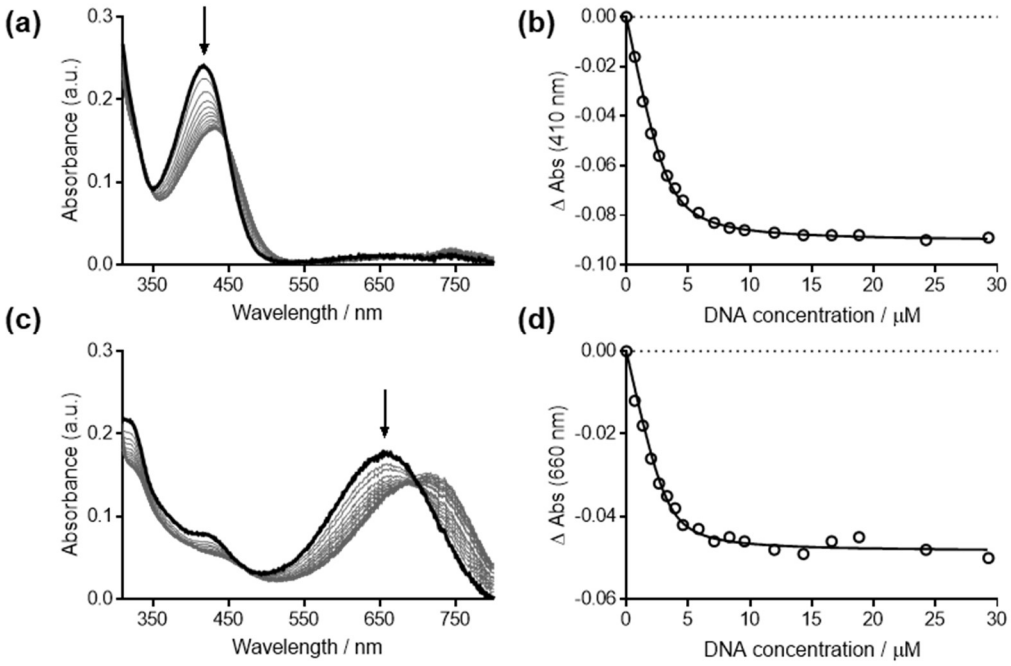


Figure 5.20: UV/visible titration studies of ligands **78o** and **78c** with hybrid telo23-K^+. (a),(c): titrations of 10 μM (a) **78o** and (c) **78c** with G4 DNA; (b),(d): extracted binding isotherms.

Overall, the UV visible titration studies confirm all DTE interact with G4 DNA occurs as suggested by the FRET assay, based on the shapes of the obtained isotherms and the approximate K_a values extracted from the independent-and-equivalent sites treatment of the data. Unfortunately, the results do not show a large affinity difference between the open and closed ligand forms which might be required to exert appreciable effects on G4 structures *in vivo*. Whilst this is perhaps expected on the basis of the FRET assay, which suggests the difference in activity between the isomers is modest, it is surprising in light of the NMR experiments which indicate different ligand binding modes that would reasonably be expected to occur with different affinities. However, the model used to extract binding constants from the UV data relies on the assumption that the multiple ligand binding events are equivalent and occur statistically rather than co-operatively. This assumption is likely to be inaccurate, and the K_a values obtained should therefore be interpreted with caution. Further isothermal experiments to compare the relative binding affinity of the open and closed isomers, for G4 such as competition dialysis studies, fluorescence indicator displacement assays or isothermal titration calorimetry should be employed to further interrogate the possible differences ability of the two ligand forms to recognise G4 structures, but unfortunately these could not be pursued in the timeframe of this project.

5.6 *In vitro* assays

Despite the apparent similarly in affinity between ligand **77o** and **77c** suggested by the UV/visible studies, the differences in the interaction modes of these ligands with G4 suggests the different photoisomeric forms of these ligands may have potential to perturb G4 formation in different ways in biological systems. At this stage of the project, a detailed investigation of the biological activity of the compounds was not possible due to time constraints, but the toxicity of the two forms of ligand **77o/77c** was measured in HeLa cells by Sadiyah Sheikh to provide an initial assessment of the potential of the **77o/77c** ligand couple as a photopharmacological anticancer agent (Figure 5.21). Interestingly, alamarBlue cell viability assays carried out in HeLa cells exposed to the two ligand forms, **77o** and **77c**, over 72 hours showed a two-fold difference in cytotoxicity between the two ligands. The closed form (EC_{50} = 10 μM) appearing more potent than the open form (EC_{50} = 22 μM). Whilst this preliminary result certainly does not provide a definitive link between the G4 binding properties of **77o** and **77c**, it does suggest that the bioactivity of these derivatives warrants further investigation and potential photopharmacological agents.

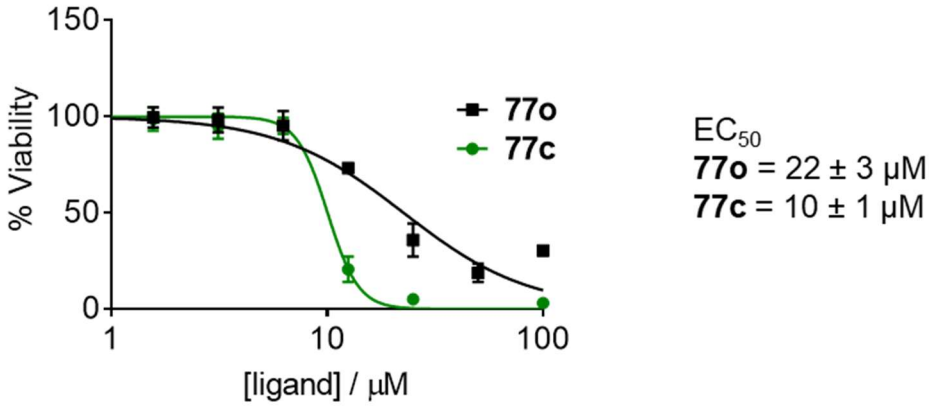


Figure 5.21: Dose-response curves for **77o** and **77c** against HeLa cells at 72 h exposure, measured by the alamarBlue cell viability assay.

5.7 Chapter conclusions

This work reported in this chapter aimed to employ the DTE photoswitch to address the key limitation of the previous stiff-stilbene system, seeking to achieve a G4 ligand for which activity can be controlled with light in a fully bi-directional manner.

From the CD and NMR experiments, it can be concluded that the binding mode of DTE ligands can be reversibly manipulated *in situ* by controlling the structure of the ligand by

photoisomerisation. This represents an advantage over the stiff-stilbene system reported in the previous chapters where the ligand photochemistry is irreversible. Whilst the change in binding mode observed for the antiparallel (telo22-Na^+) G4 appears to take place without perturbing the underlying G4 structure, both CD and NMR suggest that the structure of the G-tetrads is affected by the photoisomeric state of the **67o/67c** couple. Whilst the difference in binding modes did not appear to correlate with a significant difference in binding affinities based on the results of the UV/visible titration experiments, the different photoisomers of **77o/77c** did exert differing toxicity to cancer cells, suggesting that photoresponsive G4 ligands may find application in photopharmacological anticancer strategies. Screening of more potent DTE ligands, such as quinolinium derivatives **78o/78c**, against a wider panel of cell lines, followed by more thorough mechanism-of-action studies, is warranted in order to develop this research avenue.

6 Conclusions and future work

The overall goal of this thesis was the design of photoresponsive G-quadruplex ligands that allow G4 stability and structure to be photoregulated in biologically-relevant conditions, towards the development of G4-targeted photopharmacology or G4-based nanodevices. In pursuit of this goal, the potential of two photoresponsive scaffolds (stiff-stilbene and dithienylethene) to serve as the basis of G4-targeting ligands has been investigated. This has involved the design and synthesis of prototype ligands derived from each scaffold, investigation of their DNA-binding properties using a suite of biophysical methods, studies into the photochemistry of the compounds and preliminary assessments of their potential biological utility.

Stiff-stilbene was first explored in attempt to build on previous work on azobenzene derivatives and deliver compounds that exert more powerful effects on G4 folding in physiologically-relevant conditions. A prototype amine-decorated compound (*E*-**41**) demonstrated good stabilisation of quadruplex DNA structures whilst remaining ineffective at stabilising duplex DNA at the same concentrations. This selectivity was confirmed through competition experiments. Interestingly, the ligand appeared to bind preferentially to the hybrid telomeric G4 in preference to the antiparallel form in both the FRET and UV assays, though further screening of additional G4 structures should be undertaken verify this interesting selectivity. Critically, the *Z*-**41** isomer was demonstrated to be ineffective at stabilising G4 DNA, providing an initial validation of the hypothesis that G4 ligand activity could ultimately be controlled by stiff-stilbene isomerisation.

Unfortunately however, neither isomer perturbed the folded G4 structures in physiologically-relevant buffers. Moreover, whilst the compounds displayed anticancer and antiparasitic activity in cellular assays, the activity did not correlate with their G4 binding properties and no selectivity between tumoral and non-tumoral cells was observed. This suggested that the G4 affinity of these compounds was likely too low to result in G4-related mechanisms of therapeutic activity. Efforts to optimise the G4 binding properties by varying the substitution pattern, linker unit and amine moieties were ultimately unsuccessful and demonstrated only very limited effects on the activity of the compounds.

This initial setback was overcome by an alternative ligand design strategy, in which the flexible amine-terminated chains were replaced with *N*-methylpyridinium moieties (Chapter 3). This led to significantly improved G4 binding properties. In particular, compound *E*-**56** demonstrated particularly high (ΔT_m = 21 °C) stabilisation of G4 DNA structures at low ligand concentration, with good discrimination against duplex DNA in a competition assay.

Pleasingly, the difference in activity between the *E* and *Z* isomers was retained, with the isomer *Z*-**56** being comparatively less effective at stabilising G4 DNA in a similar manner to the first-generation compounds.

Unlike the amine-decorated compounds, the *N*-methylpyridinium compounds were found to exert significant effects on G4 folding equilibria. These effects were found to be most striking against the telo23-Na^+ sequence in sodium buffer, with *Z*-**56** stabilising an antiparallel structure, whilst the corresponding *E*-**56** isomer induced the unfolding of the G4. Evidence of this effect was observed by CD spectroscopy, NMR spectroscopy and investigated further by molecular dynamics studies. Meanwhile, all ligands were found to maintain G4 structures in cell-like potassium-rich conditions, although different binding modes were observed by CD and NMR spectroscopy, with ligand *E*-**56** appearing to target the G4 grooves rather than the G-tetrads.

Interestingly, the bioactivity of the compounds appeared to correlate with G4 binding ability, though further work must be undertaken to obtain stronger evidence that the mechanism of action truly involves G4 targeting. In particular, the compounds should be screened against a much wider panel of cell lines than was possible in the current project, to ensure the selectivity for tumoral and non-tumoral cells at long exposure times is a general effect. Furthermore, the mechanism of action of the compounds should be elicited. For example, the telomerase repeat amplification protocol (TRAP) assay could be used to prove that the compounds inhibit telomerase through G4 binding. Meanwhile, gene expression assays should examine whether transcription of G4-containing sequences is selectively modulated by the compounds, and which specific genes are affected. Positive results in this area would then justify further optimisation of the ligand scaffold for pharmacological applications.

Nonetheless, these present results reveal the lead compound to be 29-fold selective for a cancerous cell model over non-tumoral cells (EC_{50} HeLa = 62 nM), suggesting a promising therapeutic window for this molecule. The compounds also demonstrated potential as antiparasitic agents, with selectivity of up to 230-fold for *T. brucei* over non-tumoral human cells. Most importantly, towards the goal of photopharmacology, *E*-**56** was found to be 85-fold more active than the *Z*-**56** counterpart, suggesting the potential to exploit stiff-stilbene compounds as photoresponsive drugs.

Towards this end, the photochemistry of the *E/Z*-**56** pair was investigated (Chapter 4). Though *Z*-**56** was found to photoisomerise to *E*-**56** in DMSO, in aqueous conditions irradiation unfortunately led to photooxidation meaning that the reversible control of bioactivity could not be pursued further for this compound. Nonetheless, ligand *E*-**56** could be deployed as a photoresponsive fuel to regulate the formation of G4 in sodium-rich conditions, based on the ligand-prompted unfolding of this structure which is subsequently reversed upon

photoirradiation. Whilst applications of this system can be envisaged, such as the phototriggered release of a G4 to sequester a protein target, the lack of full photoreversibility would likely limit the utility of such a system.

Towards addressing this limitation, DTE derivatives were later investigated (Chapter 5). Though these molecules have previously only been studied as duplex DNA binders, it was found that *N*-methylpyridinium **77o/77c** and *N*-methylquinolinium compounds **78o/78c** induce selective stabilisation of G4 DNA in thermal melting assays, although they were not as powerful as the best stiff-stilbene compounds. The quinolinium derivatives were observed to be more effective, suggesting that the increased ligand surface area is beneficial for G4 recognition in these compounds. Unfortunately, the difference in activity between the two photoisomers was not as marked as between the stilbene *E/Z* forms, but the closed ligands nonetheless induced approximately 3 °C higher thermal stabilisation in G4 than the open forms for the F21T-K^+ and FmycT.

Despite the similarities observed in the FRET assay, **77o/77c** were found to display significantly different binding modes to G4 DNA. In the case of hybrid telo23-K^+ G4, the ligand appeared to switch from a stacking (**77c**) to intercalative (**77o**) binding mode causing some disruption of the G-tetrad network, evidenced by disappearance of imino resonances in the NMR spectrum and attenuation of ellipticity in the CD spectrum. Meanwhile the open isomer appeared to associate with the capping residues of the lateral loops of antiparallel telo22-Na^+ whilst the closed isomer stacked with the terminal G-tetrads.

Pleasingly, in contrast to the stiff-stilbene ligands, both **77o/77c** and **78o/78c** could be isomerised between open and closed forms repeatedly in DNA buffer, though this was found to be most efficient for the **77o** analogue. Notably, the photocyclisation of this derivative was shown to be accelerated by G4 DNA, the first example of a G4-catalysed photoreaction. This interesting result should be explored further. In particular, it will be important to screen more DNA sequences to assess whether this effect is specific to G-quadruplexes.

The photoresponsive control of ligand binding mode raises the possibility that these ligands may be used as supramolecular tools to modulate the behaviour of G4-based functional systems. In particular, it should be investigated whether the change in binding mode of the ligand may regulate the ability of G4 to bind to a second molecule, such as thrombin, and therefore used as a trigger for G4-mediated processes or networks.

Finally, the toxicity of **77o/77c** HeLa cells was examined, revealing a 2-fold difference in cytotoxicity between the isomers. Though the effective concentrations matched the approximate G4 affinity, and the slightly more active G4 stabilising isomer (**77c**) was the more effective cytotoxic agent, there is not yet sufficient evidence to confidently ascribe G4-targeting as the mode of action. Further studies, including screening of more potent DTE G4 ligands

such as quinolinium derivatives **78o/78c,** as well as further analogues, must be undertaken to consider the generality of the approach. Nonetheless, this preliminary result suggests the utility of the photoresponsive ligand as a photopharmacological lead.

In conclusion, two photoresponsive G4 ligand sets have been identified with together offer complementary means of photocontrol over G4/ligand assemblies. The ligands are sufficiently potent to exert effects in physiologically-relevant conditions and offer a means of control of G4 ligand activity to complement the existing toolbox of G4-targeting molecules. It is hoped that this work demonstrates the potential of G4 targeting photoswitches in photopharmacology and nanotechnology, and that the reported research will generate further interest towards the development of photoresponsive G4/ligand assemblies with novel applications.

7 Experimental

7.1 Ligand preparation

For biophysical and cellular assays, stock ligand solutions were prepared at 10 mM concentration in DMSO-d_6 and stored at -20 °C. Compounds were diluted in appropriate buffer immediately prior to use.

7.2 Oligonucleotides

All oligonucleotides used were purchased from Eurogentec (Belgium), purified by HPLC and delivered dry. Oligonucleotide concentrations were determined by UV-absorbance using a NanoDrop 2000 Spectrophotometer from Thermo Scientific through the relationship:

$$\text{concentration} = A_{260} / \varepsilon_{260}$$

Where A_{260} is the absorbance of the sample at 260 nm at 1 cm path length and ε_{260} is the molar extinction coefficient at 260 nm. The sequences and respective extinction coefficients are given below:

DNA model	Sequence	ε_{260} / M^{-1} cm^{-1}
F21T	5'-FAM-GGGTTAGGGTTAGGGTTAGGG-TAMRA-3'	268300
FmycT	5'-FAM-TTGAGGGTGGGTAGGGTGGGTAA-TAMRA-3'	290100
Febr1T	5'-FAM-GGGCAGGGGGTGATGGGGAGGAGCCAGGG-TAMRA-3'	349000
F10T	5'-FAM-TATAGCTATA-HEG-TATAGCTATA-TAMRA-3'	263600
telo23	5'-TAGGGTTAGGGTTAGGGTTAGGG-3'	236500
telo22	5'-AGGGTTAGGGTTAGGGTTAGGG-3'	228500
Pu27	5'-TGGGGAGGGTGGGGAGGGTGGGGAAGG-3'	279900
ds26	5'-CAATCGGATCGAATTCGATCCGATTG-3'	253200

FAM = 6-carboxyfluorescein;

TAMRA = 6-carboxy-tetramethylrhodamine;

HEG = $[(-CH_2CH_2O-)_6]$

Stock solutions of all sequences were prepared in HPLC-grade H_2O and stored at -20 °C. Prior to use, solutions were thawed, diluted in appropriate buffer, and annealed by heating for 2 minutes at 90 °C and then placing immediately into ice.

7.3 Biophysical studies

7.3.1 FRET thermal melting assays

Fluorescence resonance energy transfer (FRET) melting assays were performed according to the procedure reported by De Cian and co-workers[108] on a Roche LightCycler 480 or a Stratagene Mx3005P qPCR instrument. The assay was performed in 96-well plate format. The final volume of solution in each well was 25 µL and contained the appropriate oligonucleotide (200 nM) in 10 mM lithium cacodylate buffer (pH 7.2) plus the following salts depending on the sequence:

DNA model	FRET buffer additives
F21T-K^+	10 mM KCl, 90 mM LiCl
F21T-Na^+	100 mM NaCl
FmycT	1 mM KCl, 99 mM LiCl
Febr1T	10 mM KCl, 90 mM LiCl
F10T	10 mM KCl, 90 mM LiCl

Final ligand concentrations were either 1 µM, 2 µM, 5 µM or 10 µM and prepared from an intermediate stock of 50 µM in HPLC-grade H_2O (0.5% DMSO). Each sample was tested in duplicate on the same plate, and each plate was repeated in at least duplicate. Samples were heated at 1 °C increments and allowed to equilibrate for one minute at each temperature before measuring FAM fluorescence (λ_{ex} = 492 nm, λ_{em} 516 nm). Appropriate negative controls (DNA in the absence of ligand) were carried out for each sample set to allow calculation of the ΔT_m. The melting temperatures were determined from the normalised curves where % fluorescence = 0.5.

7.3.2 Circular dichroism titrations

Circular dichroism titrations were recorded using a Jasco J-810 spectrometer fitted with a Peltier temperature controller. Measurements were taken at 20 °C in a quartz cuvette with a path length of 5 mm, at a 100 nm / min scanning speed at 1 nm intervals. The CD spectra were recorded between 800 and 200 nm, and baseline corrected for the buffer used. The oligonucleotide sequences used were: telo23, telo22 and Pu27. The oligonucleotides were at

a concentration of 4.2 μM in 1 mL buffer. The buffer used was either sodium (telo22) or potassium (telo23, Pu27) phosphate (100 mM, pH 7.4). The ligand was added by aliquot from a 1 mM stock solution in the appropriate buffer (10% DMSO). The reported spectrum for each sample represents the average of 3 scans. Data processing was carried out using Prism 7 with a second-order smoothing polynomial applied to all spectra. Observed ellipticities (in mdeg) were converted to molar ellipticity [θ] = deg cm^2 $dmol^{-1}$.

7.3.3 Circular dichroism kinetics

CD kinetic studies of the conformational switch of telo23-Na^+ with ligand *E*-**56** were undertaken by monitoring the ellipticity at 273 nm with time following the addition of ligand. *E*-**56** (10 equiv.) was added to the bottom of an empty cuvette (5 mm path length) in 4 μL DMSO. Following stabilisation of the background CD signal, telo23 (1 equiv., 1 mL of 4.22 μM solution in 100 mM sodium phosphate buffer, pH 7.4) was injected rapidly to facilitate mixing and the ellipticity at 273 nm monitored for 2500 sec. Three independent repeats were conducted and the change in ellipticity at 273 nm plotted as a function of time. Fitting to a single exponential function ($\theta = a - b*\exp(-t/\tau)$) was carried out using Prism 7 software to determine the characteristic folding time (τ).

7.3.4 UV/visible spectroscopy

UV spectra were recorded on a Thermo Scientific BIOMATE 3S UV/visible spectrophotometer at ambient temperature. Measurements were taken in a 3 mL quartz cuvette with a path length of 10 mm. The UV/visible spectra were recorded between 800 nm and 200 nm and baseline corrected for the buffer used.

7.3.5 UV/visible titrations

The concentration of ligand was fixed at 10 μM in a constant volume of 1.5 mL buffer. The buffer used was either sodium (telo22) or potassium (telo23) phosphate (100 mM, pH 7.4). During the titration, aliquots of sample were removed and replaced with aliquots of oligonucleotide to give the required titration points (from a 100 μM stock solution in appropriate buffer containing also 10 μM ligand to maintain constant ligand concentration throughout the titration). The oligonucleotide stock solution was annealed by heating to 90 °C for 2 minutes and then cooling on ice prior to the addition of ligand (to avoid annealing in the presence of ligand). Following addition of each DNA aliquot, the solution was mixed thoroughly, and the UV/visible spectrum was acquired immediately. Data were fit to the appropriate binding models (discussed in the relevant chapters) using non-linear regression in Prism 7 software to provide estimates of the association constants.

7.3.6 NMR spectroscopy of G-quadruplex ligand complexes

^{1}H NMR spectra of G-quadruplex sequences were recorded at 298 K using a 600 MHz Varian VNMRS spectrometer equipped with a triple resonance cryogenically cooled probe head. The oligonucleotide sequences used were telo23 and telo22. Samples of oligonucleotide were dissolved in 90% H_2O/10% D_2O containing either 25 mM sodium phosphate (pH = 7.0) and 70 mM sodium chloride (for telo22) or 20 mM potassium phosphate (pH = 7.0) and 70 mM potassium chloride (for telo23). All experiments employed sculpted excitation water suppression. NOESY spectra were acquired with a 300 ms mixing time. The final NMR samples contained 600 µL of 185 µM oligonucleotide (for 1D experiments) or 300 µM oligonucleotide (for 2D NOESY experiments). Aliquots of ligand (10 mM in DMSO-d_6) were added to yield the desired titration points, the sample was mixed thoroughly, and NMR spectra were recorded immediately. Data were processed using MestReNova software (version 11.0.2). Resonances were assigned from data provided in the literature by Wang and Patel (telo22)[27] and by Patel *et al.* (telo23).[37]

7.4 Photoirradiation experiments

7.4.1 Stiff-stilbene ligands (Chapter 4)

The 800 nm fundamental output of a Ti:Sapphire ultrafast amplifier (Libra, Coherent) was used to generate 397 nm light *via* second harmonic generation. The measured beam spot size (diameter) at the sample region was 0.4 cm and the energy used to irradiate solutions was 30 mW. Experiments involving ligand in the absence of telo23 were conducted using 1.5 mL solution in a 3 mL quartz cuvette with a path length of 10 mm equipped with a magnetic stirrer. The concentration of the ligand was 1 mM in DMSO-d_6 or 10 µM in sodium phosphate buffer (100 mM, pH 7.4). The photoreaction was followed using NMR or UV/visible spectroscopy. Experiments involving telo23 were conducted using 1 mL solution in a 2 mL quartz cuvette with a path length of 5 mm. The concentration of DNA was 4.2 µM in sodium phosphate buffer (100 mm, pH 7.4). The effect of photoirradiation on G4 topology was monitored by circular dichroism spectroscopy. For experiments involving the repeated addition of *E*-**56**, 10 equiv. of ligand were added by aliquot (82 µL) of 600 µM *E*-**56** solution in water. The system was allowed to equilibrate for 20 min at 20 °C following addition of ligand before the CD spectrum was recorded. Photoirradiation of NMR samples was conducted by transferring the sample to a 2 mL quartz cuvette with a path length of 5 mm equipped with a magnetic stirrer, to ensure even irradiation. Samples were then transferred back to the NMR tube for analysis.

7.4.2 Dithienylethene ligands (Chapter 5)

For photoirradiation experiments conducted in DMSO-d_6 in the absence of DNA, samples of ligand were prepared at 10 mM concentration and irradiated in an NMR tube or glass screw-cap vial with a 36 W UV-A lamp (320−400 nm) of domestic use (PL-L, approximate length 30cm, Philips). The isomerisation was followed by NMR spectroscopy.

Experiments in buffer were followed by UV/visible spectroscopy at a concentration of 50 μM. Samples were irradiated using monochromatic light at 450 nm (for photocyclisation) and 635 nm (for photocycloreversion). Samples (2 mL volume) were irradiated in a 3 mL quartz cuvette containing a magnetic stirrer using collimated laser diode modules (ThorLabs, CPS450 and CPS635, 4.5 mW, elliptical beam) and the UV/visible spectra recorded at appropriate time points. The degree of conversion was estimated by comparison to a reference spectrum of the photostationary state achieved in DMSO, the composition of which was determined by NMR (*vide supra*) and then diluted to 50 μM in appropriate buffer for measurement of the reference UV/visible spectra. For photoirradiation experiments conducted in the presence of oligonucleotides and followed by NMR spectroscopy, to ensure even irradiation of the sample the solution (600 μL) was transferred to a 2 mL quartz cuvette with a 5 mm path length equipped with a magnetic stirrer and irradiated at the appropriate wavelength for the time specified, then transferred back to the NMR tube for analysis.

7.5 Computational studies

7.5.1 Density functional theory calculations

Undertaken by Marta Duchi

The ground state structures of ligands *E*-**56**, *Z*-**56** and *E*-**57** were calculated using Gaussian 09 at the B3LYP/6-31G level of theory.

7.5.2 Molecular dynamics simulations

Undertaken by Dr Susanta Haldar

7.5.2.1 Standard (unbiased) molecular dynamics simulations

Docking calculations were performed using AutoDock Vina with the DNA structure kept fixed in its original solution conformation throughout the docking procedures.[225] The highest affinity binding pose was then submitted to long microsecond molecular dynamics simulations in order to predict the stability of the binding pose.

All the simulations were performed using the Gromacs-5.0 software package.[226] The recently introduced parm-BSC1 force field was used for the DNA parameterisation. For the ligand, the General Amber Force Field was used to generate parameters.[227,228] The charges were calculated using the restrained electrostatic potential (RESP) fitting procedure.[229] The RESP fit was performed onto a grid of electrostatic potential points calculated at the HF/6-31G(d) level as recommended by the force-field designers and recent literature.[228,229]

To start the simulations, each docked ligand/DNA complex was solvated in a cubic box with the dimension of 74 x 74 x 74 $Å^3$ along with 12982 TIP3P explicit water molecules.[230] An extra 19 Na^+ ions were added to neutralise the system.[231] The complexes were minimised prior to the equilibration and production run as follows: the minimisation of the solute hydrogen atoms on the DNA and the ligand was followed by the minimisation of the counterions and the water molecules in the box. In the next step, the DNA backbone along with the all the heavy atoms on the ligand were kept frozen, and the solvent molecules with counterions were allowed to move during a 50 ps MD run, to relax the density of the whole system. In the next step the nucleobases were relaxed in several minimisation runs with decreasing force constants applied to the DNA backbone atoms, however, a few phosphate atoms were kept restrained with a force constant of 2.39 $kcal.mol^{-1}$ $Å^{-2}$. After the full relaxation, the system was slowly heated to the room temperature to 300K using V-rescale thermostat with a coupling constant of 0.5 ps employing an NVT (constant-temperature, constant-volume) ensemble.[232] As the system reached the temperature of interest, the equilibration simulation was performed for 10000 ps (10 ns) using an NPT ensemble with Berendsen thermostat and Berendsen barostat, and 0.5 ps was used again as the coupling constant for both temperature and pressure, respectively.[233] Finally, the production run was set for 1000000000 ps (1μs) using Nose-Hoover thermostat[234] and Parrinello-Rahman barostat[235] with the same coupling constant as previously taken in the equilibration simulation in the NPT ensemble. All the simulations were carried out under the periodic boundary conditions. The particle-mesh Ewald method was used to calculate the electrostatic interactions with in a cut-off of 10 Å.[236] The same cut-off was used for Lennard-Jones interactions. All simulations were performed with a 1.0 fs time step.

7.5.2.2 Well-tempered metadynamics

The plumed 2.3 plugin was used to carry out the simulation with the Gromacs-5.0.7 code.[237] The bias potential was calculated according to the WTMetaD scheme as follows:

$$V(s,t) = \sum_{t'=0,\tau_G,2\tau_{G,\ldots}}^{t'<t} \omega\tau_G \, e^{-V(s(q(t'),t'))/\Delta T} \, e^{-\sum_{i=1}^{2}[(s_i(q)-s_i\left(q(t')\right))^2/2\sigma_i^2]}$$

where the deposition rate, ω, and deposition stride, τG, of the Gaussian hills were set to 0.358 kcal.mol^{-1}·ps^{-1} and 1.0 ps, respectively. The bias factor (T + ΔT)/T was set to 15, and the final free energy surface (FES) was calculated as follows:

$$F(s,t) = -\frac{T+\Delta T}{\Delta T}(V(s,t) - C(T))$$

where the V(s,t) is the bias potential added to the Collective Variables (CV) used and the T represents the simulation temperature. ∆T is the difference between the temperature of the CV and the simulation temperature. The bias potential is grown as the sum of the Gaussian hills deposited along the chosen CV space and finally the sampling of particular CV space can be controlled with the tuning of the ∆T parameter.

7.6 Cellular and parasitic studies

7.6.1 Cell and parasite culture

Conducted with assistance from Dr Pablo Peñalver

MRC-5 cells (human lung fibroblasts) were grown in monolayer (37°C, 5% CO_2 and 100% humidity) in DMEM medium (1 g/L glucose), supplemented with 10% heat-inactivated Fetal Bovine Serum, 2 mM L-glutamine, 100 U/mL penicillin and 100 mg/mL streptomycin. HeLa cells were maintained at 37 °C and 5 % CO_2 in high glucose DMEM (4.5 g/L) supplemented with 10% hiFBS, 100 U/ml penicillin, 100 mg/ml streptomycin, 2 mM L-glutamine and non-essential amino acids (1X). Cells were cultured according to ATCC recommendations and were used for the experiments while in the exponential growth phase.

T. brucei (bloodstream forms, “single marker” S427 (S16)) were cultured at 37 °C, 5 % CO_2 in HMI-9 medium supplemented with 10% heat-inactivated fetal bovine serum (hiFBS, Invitrogen). *L. major* promastigotes (MHOM/IL/80/Friedlin) were cultured at 28 °C, 5 % CO_2 in modified RPMI-1640 medium (Invitrogen, Carlsbad, CA) with 10% hiFBS. Parasites were split every other day and maintained in their experimental growth phase (below 2 million parasites per mL).

7.6.2 Cytotoxicity

Conducted with assistance from Dr Pablo Peñalver

Cytotoxicity was measured through the alamarBlue assay (ThermoFisher). 5x10^3 MRC-5 or HeLa cells were seeded in 96-wells plates (100 µL/well) in the presence of increasing

concentrations of ligands. After 72h / 7d incubation at 37 °C, 20 µL of alamarBlue solution (110 ng/ml) was added to each well and cells were reincubated for 1 hr at 37 °C and then analysed by fluorescence (λ_{ex} = 550 nm, λ_{em} 590 nm) using an Infinite F200 plate reader (TECAN Austria, GmbH). The results are expressed as the concentration of compound that reduces cell growth by 50% versus untreated control cells (EC_{50}) using Prism software to fit the data to a sigmoidal curve. Data are presented as the average of two independent measurements all conducted in duplicate conditions.

7.6.3 Parasitic toxicity

Conducted with assistance from Dr Pablo Peñalver

$1x10^3$ BSF *T. brucei* were incubated in 96-wells plates alone or in the presence of increasing concentrations of compounds for 72 hrs (5% CO_2, 37°C). After 72 h of incubation at 37 °C, 20 µL of alamarBlue solution (110 ng/mL) was added to each well and cells were reincubated for 4 hrs at 37 °C. Then, 50 µL of 3% SDS were added to each well. The plate was incubated at 37 °C for an extra hour and then analysed by fluorescence using an Infinite F200 plate reader (TECAN Austria, GmbH) as above. Data are presented as the average of two independent measurements all conducted in duplicate conditions.

$4×10^6$/mL *L. major* promastigotes (MHOM/IL/80/Friedlin) were incubated for 72 h at 28 °C in 96-well plates (50 µL/well) in the presence of increasing concentration of compounds. Proliferation was determined using the MTT-based assay (Sigma-Aldrich). After the 72 h incubation, 10 µL of MTT (5 mg/mL) was added to each well and parasites were reincubated for 4 h at 28 °C. Then, 50 µL of 20% SDS was added to each well. The plate was then incubated at 37 °C for 4-16 h and analysed using an Infinite F200 plate reader (TECAN Austria, GmbH). The absorbance was measured at a wavelength of 540 nm. Data are presented as the average of two independent measurements all conducted in duplicate conditions.

7.6.3 Microscopy

Undertaken by Dr Pablo Peñalver

HeLa cells (20,000) and *T. brucei* parasites (10,000,000) were incubated with 0.5 and 1 µM ligand *E*-**57**, respectively, in 0.5 mL of each respective medium (without FBS) for 30 and 120 min at 37 °C and 100% humidity. Deep Red Mitotracker (400 nM) was also added to the incubation well for the last 30 minutes.

Parasites were fixed with cold paraformaldehyde 4% for 30 min, washed twice in cold phosphate buffered saline (PBS), and processed by microscope observation. Cells were

washed 5 times with room temperature PBS and fixed with paraformaldehyde 2% for 20 minutes. Two extra PBS washings and an ethanol immersion of the cover slide were also necessary prior to sample processing for microscopy. ProLong DAPI (3-4 µL) was used as mounting medium in both cases.

Images were acquired using a widefield Olympus ix81 microscope. Excitation was performed with the 350-450, 492-518 and 572-623 nm filters for DAPI, ligand **3** and Deep Red Mitotracker, respectively. The emission of ligand *E*-**57** was detected between 510 and 550 nm with a narrow filter, whereas a triple filter (437-474, 508-550 and 595-670 nm) was used to detect the fluorescence emission of both DAPI and Mitotracker.

The images were deconvoluted using Huygens Professional image processing software from Scientific Volume Imaging. The merges of the images were made with Fiji software.

7.7 Synthetic procedures and compound characterisation

General experimental

Chemicals were purchased and used without further purification. Dry solvents were obtained by distillation using standard procedures, or by passage through a column of anhydrous alumina using equipment from Anhydrous Engineering (University of Bristol) based on the Grubbs' design.[238] Reactions requiring anhydrous conditions were performed under N_2; glassware and needles were either flame dried immediately prior to use, or placed in an oven (150 °C) for at least 2 h and allowed to cool in a desiccator or under reduced pressure. Liquid reagents, solutions or solvents were added via syringe through rubber septa; solid reagents were added via Schlenk type adapters. Reactions were monitored by TLC on Kieselgel $60F_{254}$ (Merck), with UV light (254 nm) detection and by staining with basic potassium permanganate solution. Flash column chromatography was performed according to the method reported Still and co-workers using silica gel [Merck, 230−400 mesh (40−63 μm)].[239] Solvents for flash column chromatography (FCC) and thin layer chromatography (TLC) are listed in volume:volume percentages. Extracts were concentrated *in vacuo* using both a Heidolph HeiVAP Advantage rotary evaporator (bath temperatures up to 50 °C) at a pressure of 15 mmHg (diaphragm pump) or 0.1 mmHg (oil pump), as appropriate, and a high vacuum line at room temperature.

Preparative HPLC was performed on a Grace Discovery Sciences Reveleris Prep System with a Phenomenex Luna 5 μm C18(2) 100 Å AXIA packed (250 × 21.2 mm) column. For purification, the instrument was set to monitor the ELSD signal as well as 220 nm, 254 nm and 280 nm on the UV detector. Flow rates were 14 mL / min. Water soluble compounds were freeze dried on a Lyotrap Plus (LTE Scientific LTD).

^{1}H NMR and ^{13}C NMR spectra were measured at 25°C in the solvent specified with Varian or Bruker spectrometers operating at field strengths listed. Chemical shifts are quoted in parts per million with spectra referenced to the residual solvent peaks. Multiplicities are abbreviated as: br (broad), s (singlet), d (doublet), t (triplet), q (quartet), p (pentet), m (multiplet) and app. (apparent) or combinations thereof. Assignments of ^{1}H NMR and ^{13}C NMR signals were made where possible, using COSY, HSQC and HMBC experiments. Mass spectra were obtained by the University of Bristol mass spectrometry service by electrospray ionisation (ESI), electron ionisation (EI) or matrix assisted laser desorption ionisation (MALDI) modes. Infra-red spectra were recorded in the range 4000-400 cm^{-1} on a Perkin Elmer Spectrum either as neat films or solids compressed onto a diamond window.

General procedure A (for formation of anilines)

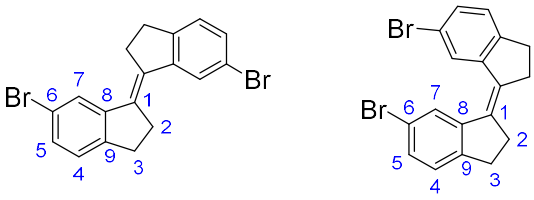
The appropriate bis-bromide (1 equiv.), BrettPhos Pd G3 (5 - 10 mol%) and sodium *tert*-butoxide (4 equiv.) were added to a 10 mL reaction tube containing a stirrer bar. The tube was sealed with a septum crimp cap and the contents dried under high vacuum for 30 min. The tube was then purged with nitrogen and evacuated and backfilled with nitrogen a further three times. 1,4-Dioxane (2 mL) was added followed by the appropriate amine (2.2 equiv.). The reaction mixture was then heated to 80 °C and stirred for 2 h, then cooled and concentrated *in vacuo*. The crude mixture was filtered through a plug of silica, eluting with 80:20:1 DCM:MeOH:Et_3N. The solvent was removed *in vacuo* and the residue suspended in H_2O (4 mL). Trifluoroacetic acid (200 µL) was added, the mixture swirled to dissolve the majority of the solid, and the resulting suspension filtered. The filtrate was purified by reverse phase HPLC eluting with a gradient of 5% MeCN in H_2O (0.05% TFA) to 80% MeCN in H_2O (0.05% TFA) over 30 minutes. Lyophilisation of the relevant fractions afforded the target compounds as hygroscopic solids.

General procedure B (for formation of ureas)

The appropriate bis-aniline (1 equiv.) was added to a 10 mL reaction tube containing a stirrer bar. The tube was sealed with a septum crimp cap and the contents dried under high vacuum for 30 minutes. The tube was then purged with nitrogen and evacuated and backfilled with nitrogen three times. Anhydrous dichloromethane (1 mL) was added, followed by a solution of triphosgene (0.66 equiv.) in anhydrous DCM (1 mL). The reaction was stirred at room temperature for 10 min, then triethylamine (3 equiv.) was added, followed by the appropriate amine. The reaction was stirred at room temperature for 1 h, then concentrated *in vacuo* and suspended in H_2O (4 mL). Trifluoroacetic acid (200 µL) was added, the mixture swirled to dissolve the majority of the solid, and the resulting suspension filtered. The filtrate was purified by reverse phase HPLC eluting with a gradient of 5% MeCN in H_2O (0.05% TFA) to 80% MeCN in H_2O (0.05% TFA) over 30 minutes. Lyophilisation of the relevant fractions afforded the target compounds as hygroscopic solids.

(*E*)-6,6'-dibromo-2,2',3,3'-tetrahydro-1,1'-biindenylidene, (*E*)-43

(*Z*)-6,6'-dibromo-2,2',3,3'-tetrahydro-1,1'-biindenylidene, (*Z*)-43



Zinc powder (5.90 g, 90.2 mmol) was suspended in anhydrous THF (75 mL). $TiCl_4$ (4.9 mL, 45 mmol) was added dropwise as the solution was stirred vigorously. The solution was heated to reflux and stirred for 2 h. The mixture was then cooled to r.t. and 6-bromo-1-indanone (4.76 g, 22.6 mmol) was added to the suspension. The mixture was heated to reflux and stirred for 16 h, then quenched with sat. aq. NH_4Cl solution (50 mL) and extracted with $CHCl_3$ (3 × 100 mL). The combined organic extractions were washed with H_2O (100 mL) and then concentrated *in vacuo* to approximately 30 mL, at which volume a precipitate formed. The precipitate was collected by filtration under reduced pressure and dried under vacuum to afford bromide *E*-**43** as a yellow solid (2.44 g, 55%). The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with hexane, to afford bromide *Z*-**43** as a white solid (1.00 g, 22%).

E-**43**:

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.68 (2H, d, *J* = 1.8 Hz, 7-C*H*), 7.41 (2H, dd, *J* = 8.0, 1.8 Hz, 5-C*H*), 7.33 (2H, d, *J* = 8.0 Hz, 4-C*H*), 3.13 – 3.08 (4H, m, 2-CH_2 or 3-CH_2), 3.07 – 3.02 (4H, m, 2-CH_2 or 3-CH_2).

^{13}C NMR Too insoluble for this experiment.

ν_{max} / cm^{-1} (compressed solid) 2925 (w), 2887 (w), 2850 (w), 1889 (w), 1589 (m), 1560 (m), 1464 (m), 1433 (m), 1405 (m), 1205 (m), 1174 (m) 1078 (m), 1037 (m), 879 (s), 819 (s).

EI-LRMS for $C_{18}H_{14}Br_2^+$ $[M]^+$ calcd: 389.9, found: 390.2.

Proton NMR was consistent with literature data.[178]

Z-**43**:

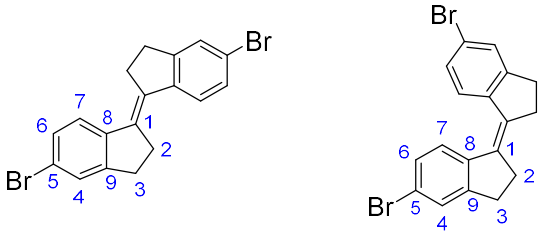
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.01 (2H, d, *J* = 1.8 Hz, 7-C*H*), 7.39 (2H, dd, *J* = 8.0, 1.8 Hz, 5-C*H*), 7.32 (2H, d, *J* = 8.0 Hz, 4-C*H*), 2.93 – 2.90 (4H, m, 2-CH_2 or 3-CH_2), 2.80 – 2.76 (4H, m, 2-CH_2 or 3-CH_2).

^{13}C NMR (126 MHz, DMSO-d_6) δ 147.6 (9-*C*), 141.8 (8-*C*), 135.2 (1-*C*), 130.2 (7-*C*H), 127.6, 125.0 (4-*C*H and 5-*C*H), 118.7 (6-*C*), 34.4 (2-*C*H or 3-*C*H), 29.6 (2-*C*H or 3-*C*H).

ν_{max} / cm^{-1} (compressed solid) 2029 (w), 2895 (w), 2871 (w), 2837 (w), 1635 (w), 1590 (m), 1559 (m), 1459 (m), 1444 (m), 1425 (m), 1264 (m), 1166 (m), 1071 (m), 803 (s).

EI-LRMS for $C_{18}H_{14}Br_2^+$ $[M]^+$ calcd: 389.9, found: 390.0.
Proton and carbon NMR were consistent with literature data.[178]

(*E*)-5,5'-dibromo-2,2',3,3'-tetrahydro-1,1'-biindenylidene, *E*-45
(Z)-5,5'-dibromo-2,2',3,3'-tetrahydro-1,1'-biindenylidene. *Z*-45



Zinc powder (3.94 g, 60.3 mmol) was suspended in anhydrous THF (30 mL). $TiCl_4$ (3.3 mL, 30 mmol) was added dropwise as the solution was stirred vigorously. The solution was heated to reflux and stirred for 2 h. The mixture was then cooled to r.t. and 5-bromo-1-indanone (3.17 g, 15.0 mmol) was added to the suspension. The mixture was heated to reflux and stirred for 30 h, then quenched with sat. aq. NH_4Cl solution (50 mL) and extracted with $CHCl_3$ (3 × 100 mL). The combined organic extractions were washed with H_2O (100 mL) and then concentrated *in vacuo* to approximately 50 mL, at which volume a precipitate formed. The precipitate was collected by filtration under reduced pressure and dried under vacuum to afford bromide *E*-**45** as an off-white solid (2.01 g, 68%). The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with hexane, to afford bromide *Z*-**45** as a white solid (550 mg, 18%).

E-**45**:

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.55 (2H, s, 4-C*H*), 7.51 (2H, d, *J* = 8.4 Hz, 7-C*H*), 7.44 (2H, d, *J* = 8.4 Hz, 6-C*H*), 3.08 (8H, app. s, 2-CH_2 and 3-CH_2).

^{13}C NMR (126 MHz, DMSO-d_6) δ 149.7 (9-*C*), 141.6 (8-*C*), 134.8 (1-*C*), 129.4 (6-*C*H), 127.9 (4-*C*H), 125.9 (7-*C*H), 120.3 (5-*C*), 31.3 (2-*C*H_2 or 3-*C*H_2), 30.3 (2-*C*H_2 or 3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 2942 (w), 2922 (m), 2877 (w), 2845 (w), 2832 (w), 1853 (w), 1743 (w), 1582 (m), 1561 (w), 1466 (s), 1450 (m), 1422 (m), 1408 (m), 1368 (w), 1346 (w), 1321 (w), 1297 (m), 1280 (m), 1250 (m), 1223 (w), 1204 (m), 1711 (s), 1116 (w), 1070 (m), 871 (s), 819 (m), 783 (s), 653 (m), 551 (s), 428 (s).

MALDI-HRMS for $C_{18}H_{14}Br_2^+$ $[M]^+$ calcd: 389.9442, found: 389.9444.

Z-**45**:

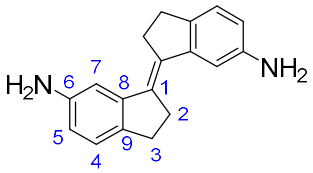
^{1}H NMR (500 MHz, DMSO-d_6) δ 7.80 (2H, d, *J* = 8.4 Hz, 7-C*H*), 7.54 (2H, d, *J* = 2.0 Hz, 4-C*H*), 7.34 (2H, dd, *J* = 8.4, 2.0 Hz, 6-C*H*), 2.99 – 2.92 (4H, m, 3-CH_2), 2.85 – 2.66 (4H, m, 2-CH_2)

^{13}C NMR (126 MHz, DMSO-d_6) δ 150.9 (9-*C*), 138.9 (8-*C*), 134.5 (1-*C*), 128.8 (6-*C*H), 128.3 (4-*C*H), 124.3 (7-*C*H), 120.2 (5-*C*), 34.3 (2-*C*H_2), 29.8 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 2938 (m), 2922 (w), 2899 (w), 2870 (m), 2834 (w), 1638 (w), 1583 (w), 1564 (w), 1467 (s), 1460 (s), 1442 (m), 1426 (m), 1411 (s), 1325 (w), 1302 (w), 1194 (w), 1170 (s), 1099 (m), 1063 (s), 1000 (m), 872 (s), 809 (s), 792 (s), 716 (m), 542 (m), 428 (s).

MALDI-LRMS for $C_{18}H_{14}Br_2^+$ $[M]^+$ calcd: 387.9457, found: 387.9455.

(*E*)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, *E*-48



Bromide *E*-**43** (398 mg, 1.03 mmol), palladium(II) acetate (28 mg, 0.11 mmol), DPPF (89 mg, 0.16 mmol) and sodium *tert*-butoxide (230 mg, 2.40 mmol) were added to a 10 mL reaction tube containing a stirrer bar. The tube was sealed with a septum crimp cap and the contents dried under high vacuum for 30 min. The tube was then purged with nitrogen and evacuated and backfilled with nitrogen a further three times. Anhydrous toluene (8 mL) was added and the solution degassed by bubbling nitrogen for five minutes. Benzophenone imine (450 µL, 2.7 mmol) was added and the reaction mixture was then heated to 90 °C and stirred for 18 h, then cooled to r.t. and partitioned between $CHCl_3$ (20 mL) and H_2O (10 mL). The aqueous layer was back-extracted with $CHCl_3$ (2 × 20 mL) and the combined organic phases dried ($MgSO_4$) filtered and concentrated *in vacuo*. The resultant oil was purified by flash silica chromatography, eluting with 1% Et_3N in DCM, to afford the imine intermediate (*E*-**46**) as a yellow solid. This was immediately re-dissolved in THF (50 mL), 2M aq. HCl (25 mL) was added and the reaction stirred at r.t. for 2 h. The mixture was then basified (pH ~ 10) by addition of a sat. aq. $NaHCO_3$ solution (110 mL), then extracted with EtOAc (3 × 50 mL). The combined organic extractions were dried ($MgSO_4$) and concentrated *in vacuo*. The residue was triturated in Et_2O (20 mL) and the resultant solid filtered and air-dried to afford the title compound as a light-brown solid (195 mg, 72%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 6.97 (2H, d, *J* = 7.9 Hz, 4-C*H*), 6.88 (2H, d, *J* = 2.0 Hz, 7-C*H*), 6.46 (2H, dd, *J* = 7.9, 2.0 Hz, 5-C*H*), 4.89 (4H, s, NH_2), 3.03 – 2.98 (4H, m, 2-CH_2 or 3-CH_2), 2.91 – 2.87 (4H, m, 2-CH_2 or 3-CH_2).

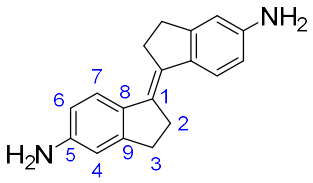
^{13}C NMR (126 MHz, DMSO-d_6) δ 147.2 (9-*C*), 143.4 (8-*C*), 134.7 (1-*C* or 6-*C*), 134.1 (1-*C* or 6-*C*), 124.8 (4-*C*H), 113.6 (5-*C*H), 109.8 (7-*C*H), 31.9 (2-*C*H_2 or 3-*C*H_2), 29.5 (2-*C*H_2 or 3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3406 (w), 3328 (w), 2847 (w), 1602 (m), 1578 (m), 1485 (s), 1450 (m), 1324 (s), 1308 (m), 1294 (w), 1281 (m), 1218 (m), 1203 (m), 1164 (w), 1105 (w), 1408 (w), 967 (w), 827 (s), 610 (s).

ESI-LRMS for $C_{18}H_{19}N_2^+$ $[M+H]^+$ calcd: 263.2, found: 263.2.

Proton and carbon NMR were consistent with literature data.[178]

(*E*)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diamine, *E*-49



Bromide *E*-**45** (752 mg, 1.92 mmol), palladium(II) acetate (52 mg, 0.21 mmol), DPPF (169 mg, 0.305 mmol) and sodium *tert*-butoxide (371 mg, 3.86 mmol) were added to a 50 mL reaction tube containing a stirrer bar. The tube was sealed with a septum crimp cap and the contents dried under high vacuum for 30 min. The tube was then purged with nitrogen and evacuated and backfilled with nitrogen a further three times. Anhydrous toluene (10 mL) was added and the solution degassed by bubbling nitrogen for five minutes. Benzophenone imine (0.81 mL, 4.8 mmol) was added and the reaction mixture was then heated to 90 °C and stirred for 18 h, then cooled to r.t. and partitioned between DCM (50 mL) and H_2O (50 mL). The aqueous layer was back-extracted with DCM (2 × 50 mL) and the combined organic phases dried ($MgSO_4$) filtered and concentrated *in vacuo*. The residue was purified by flash silica chromatography, eluting with DCM to 10% MeOH in DCM (1% Et_3N), to afford the imine intermediate (*E*-**47**) as an orange solid. This was immediately redissolved in THF (80 mL), 2 M aq. HCl (30 mL) was added and the reaction stirred at r.t. for 3 h. The mixture was then basified (pH ~ 8) by addition of a sat. aq. $NaHCO_3$ solution (100 mL), then extracted with EtOAc (3 × 50 mL). The combined organic extractions were dried ($MgSO_4$) and concentrated *in vacuo*. The residue was triturated in Et_2O (20 mL) and the resultant solid filtered and air-dried to afford the title compound as a light-brown solid (287 mg, 57%).

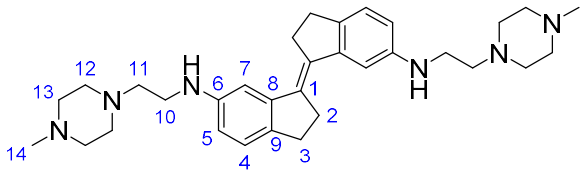
^{1}H NMR (500 MHz, DMSO-d_6) δ 6.32 (2H, d, *J* = 8.3 Hz, 7-C*H*), 5.64 (2H, d, *J* = 2.2 Hz, 4-C*H*), 5.59 (2H, dd, *J* = 8.3, 2.2 Hz, 6-C*H*), 4.23 (4H, s, NH_2), 2.03 (8H, app. s, 2-CH_2 and 3-CH_2)

^{13}C NMR (126 MHz, DMSO-d_6) δ 147.4 (9-*C* or 5-*C*), 147.3 (9-*C* or 5-*C*), 131.8 (8-*C*), 129.9 (1-*C*), 124.2 (7-*C*H), 112.6 (6-*C*H), 110.0 (4-*C*H), 31.2 (2-*C*H_2), 30.4 (3-*C*H_2)

ν_{max} / cm^{-1} (compressed solid) 3389 (m), 3192 (w), 2837 (w), 1602 (s), 1579 (w), 1488 (s), 1455 (s), 1334 (m), 1302 (m), 1269 (w), 1259 (s), 1198 (m), 854 (m), 822 (m), 795 (m), 722 (m), 733 (s), 707 (s), 698 (s), 575 (s).

ESI-HRMS for $C_{18}H_{19}N_2^+$ [M+H]$^+$ calcd: 263.1543, found: 263.1549.

(*E*)-*N*^{6},*N*$^{6'}$-bis(3-(4-methylpiperazin-1-yl)ethyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, A



. 6 TFA

Synthesised by R.L.S. Fisher, MSc student

Prepared according to General Procedure A using bromide *E*-**43** (50 mg, 0.13 mmol), BrettPhos Pd G3 (8.0 mg, 8.8 μmol), sodium *tert*-butoxide (50 mg, 0.51 mmol) and 2-(4-methylpiperazin-1-yl)ethan-1-amine (37 μL, 0.26 mmol). After lyophilisation following HPLC, the title compound was obtained as a brown solid (14 mg, 9%).

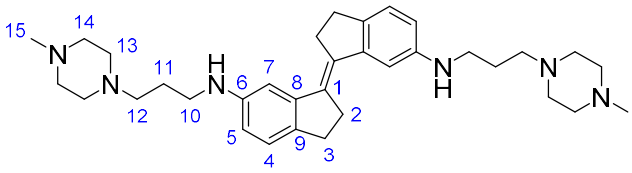
^{1}H NMR (500 MHz, D_2O): δ 7.59 (2H, d, *J* = 2.0 Hz, 7-C*H*), 7.54 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.27 (2H, dd, *J* = 8.1, 2.0 Hz, 5-C*H*), 3.67 (4H, t, *J* = 5.7 Hz, 10-CH_2), 3.65 – 3.50 (4H, br, 13-CH_2), 3.34 – 3.26 (8H, br, 12-CH_2 and 13-CH_2), 3.19 (8H, app. s, 2-CH_2 and 3-CH_2), 2.94 (6H, s, 14-CH_3), 2.92 (4H, t, *J* = 5.7 Hz, 11-CH_2), 2.74 (4H, s, 12-CH_2).

^{13}C NMR (126 MHz, D_2O): δ 162.9 (q, *J* = 35.5 Hz, CF_3*C*O), 147.9 (9-*C*), 143.9 (8-*C*), 136.2 (1-*C*), 134.6 (6-*C*), 126.5 (4-*C*H), 119.8 (5-*C*H), 117.5 (7-*C*H), 116.3 (q, *J* = 291.8 Hz, *C*F_3CO), 52.7 (13-*C*H_2), 51.6 (11-*C*H_2), 49.2 (12-*C*H_2), 46.6 (10-*C*H_2), 42.7 (14-*C*H_3), 31.6 (2-*C*H_2), 30.1 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid): 1670 (s), 1459 (m), 1411 (m), 1373 (w), 1317 (w), 1169 (s), 1124 (s), 966 (s), 830 (s), 796 (s), 721 (s).

ESI-HRMS: for $C_{32}H_{47}N_6^+$ [M+H]$^+$ calcd: 515.3857, found: 515.3856.

(*E*)-N6,N6'-bis(3-(4-methylpiperazin-1-yl)propyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, (*E*)-41 = B



. 6 TFA

Prepared according to General Procedure A using bromide *E*-**43** (49 mg, 0.13 mmol), BrettPhos Pd G3 (7.0 mg, 7.7 µmol), sodium *tert*-butoxide (49 mg, 0.51 mmol) and 3-(4-methylpiperazin-1-yl)propan-1-amine (49 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (73 mg, 47%).

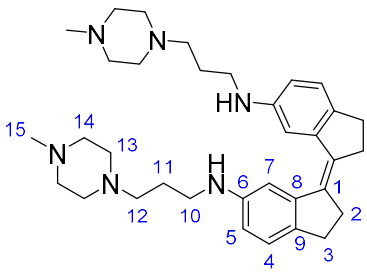
^{1}H NMR (400 MHz, D_2O) δ 7.67 (2H, d, *J* = 1.6 Hz, 7-C*H*), 7.56 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.33 (2H, dd, *J* = 8.1, 1.6 Hz, 5-C*H*), 3.79 – 3.46 (16H, br, 13-CH_2 and 14-CH_2), 3.58 (4H, t, *J* = 7.7 Hz, 10-CH_2), 3.34 (4H, t, *J* = 8.0 Hz, 12-CH_2), 3.20 (8H, app. s, 2-CH_2 and 3-CH_2), 3.02 (6H, s, 15-CH_3), 2.30 – 2.20 (4H, m, 11-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.5 Hz, CF_3CO), 149.5 (9-*C*), 144.2 (8-*C*), 136.3 (1-*C*), 133.2 (6-*C*), 126.7 (4-*C*H), 121.0 (5-*C*H), 118.0 (7-*C*H), 116.3 (q, *J* = 291.8 Hz, CF_3CO), 53.5 (12-CH_2), 50.4 (14-CH_2), 48.9 (13-CH_2), 48.1 (10-CH_2), 42.8 (15-CH_3), 31.5 (2-CH_2), 30.2 (3-CH_2), 20.6 (11-CH_2).

ν_{max} / cm^{-1} 1670 (s), 1460 (m) 1418 (m), 1176 (s), 1120 (s), 961 (m), 831 (s), 797 (s), 720 (s).

ESI-HRMS for $C_{34}H_{51}N_6^+$ $[M+H]^+$ calcd: 543.4170, found 543.4163.

(*Z*)-*N*^{6},*N*$^{6'}$-bis(3-(4-methylpiperazin-1-yl)propyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine (*Z*-41)



. 6 TFA

Prepared according to General Procedure A using bromide *Z*-**43** (59 mg, 0.15 mmol), BrettPhos Pd G3 (6.0 mg, 6.6 µmol), sodium *tert*-butoxide (58 mg, 0.61 mmol) and 3-(4-

methylpiperazin-1-yl)propan-1-amine (44 µL, 0.33 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (35 mg, 19%).

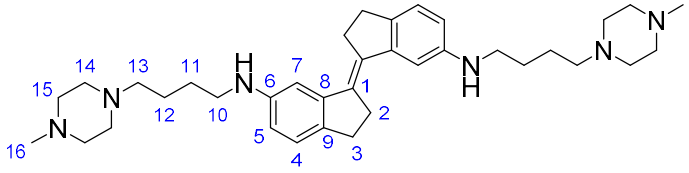
^{1}H NMR (500 MHz, D_2O) δ 7.95 (2H, s, 7-C*H*), 7.56 (2H, d, *J* = 8.0 Hz, 4-C*H*), 7.29 (2H, d, *J* = 8.0 Hz, 5-C*H*), 3.91 – 3.18 (16H, br, 13-CH_2 and 14-CH_2), 3.47 (4H, t, *J* = 7.9 Hz, 10-CH_2), 3.27 (4H, t, *J* = 8.1 Hz, 12-CH_2), 3.07 – 3.01 (4H, m, 2-CH_2 or 3-CH_2), 3.00 (s, 6H, 15-CH_3) 2.93 – 2.87 (4H, m, 2-CH_2 or 3-CH_2), 2.14 – 2.04 (4H, m, 4H, 11-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.4, CF_3CO), 149.4 (9-*C*), 141.5 (8-*C*), 136.0 (1-*C*), 133.8 (6-*C*), 127.2 (4-*C*), 120.5, (5-*C*H), 116.3 (q, *J* = 291.7 Hz, *C*F_3CO), 115.6 (7-*C*H), 53.5 (12-*C*H_2), 50.8 (14-*C*H_2), 48.9 (13-*C*H_2), 47.2 (10-*C*H_2), 42.7 (15-*C*H_3), 34.6 (2-*C*H_2 or 3-*C*H_2), 29.9 (2-*C*H_2 or 3-*C*H_2), 20.9 (11-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 1669 (s), 1439 (m), 1179 (s), 1125 (s), 839 (s), 799 (s), 723 (s).

ESI-HRMS for $C_{34}H_{51}N_6^+$ $[M+H]^+$ calcd: 543.4170, found 543.4179.

(*E*)-*N*^{6},*N*$^{6'}$-bis(3-(4-methylpiperazin-1-yl)butyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, C



. 6 TFA

Synthesised by R.L.S. Fisher, MSc student

Prepared according to General Procedure A using bromide *E*-**43** (50 mg, 0.13 mmol), BrettPhos Pd G3 (8.0 mg, 8.8 µmol), sodium *tert*-butoxide (50 mg, 0.51 mmol) and 4-(4-methylpiperazin-1-yl)butan-1-amine (45 µL, 0.26 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (51 mg, 31%).

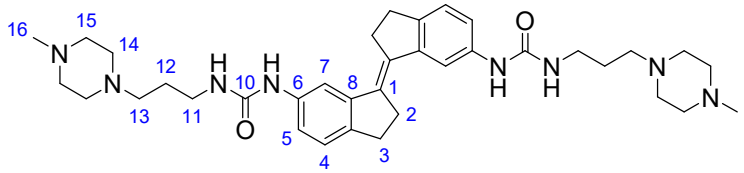
^{1}H NMR (500 MHz, D_2O): δ 7.66 (2H, d, *J* = 2.0 Hz, 7-C*H*), 7.55 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.32 (2H, dd, *J* = 8.1, 2.0 Hz, 5-C*H*), 3.73 – 3.53 (16H, br, 14-CH_2 and 15-CH_2), 3.51 (4H, t, *J* = 7.3 Hz, 10-CH_2), 3.25 (4H, t, *J* = 8.2 Hz, 13-CH_2), 3.20 (8H, app. s, 2-CH_2 and 3-CH_2), 3.00 (6H, s, 16-CH_3), 1.91 – 1.81 (8H, m, 11-CH_2 and 12-CH_2).

^{13}C NMR (126 MHz, D_2O): δ 162.9 (q, *J* = 35.5 Hz, CF_3CO), 149.3 (9-*C*), 144.1 (8-*C*), 136.3 (1-*C*), 133.4 (6-*C*), 126.6 (4-*C*H), 121.1 (5-*C*H), 118.1 (7-*C*H), 116.3 (q, *J* = 291.8 Hz, *C*F_3CO), 56.0 (13-*C*H_2), 50.8 (10-*C*H_2), 50.4 (15-*C*H_2), 48.7 (14-*C*H_2), 42.8 (16-*C*H_3), 31.5 (2-*C*H_2), 30.2 (3-*C*H_2), 22.4 (11-*C*H_2 or 12-*C*H_2), 20.8 (11-*C*H_2 or 12-*C*H_2).

ν_{max} / cm^{-1} (compressed solid): 1670 (s), 1594 (w), 1483 (m), 1451 (m), 1415 (s), 1192 (s), 1176 (s), 1114, (s), 1019 (m), 958 (s), 831 (s), 796 (s), 719 (s).

ESI-HRMS: for $C_{36}H_{55}N_6^+$ $[M+H]^+$ calcd: 571.4483, found: 571.4470.

(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(3-(3-(4-methylpiperazin-1-yl)propyl)urea), D



. 4 TFA

Prepared according to General Procedure B using bis-aniline (*E*)-**14** (30 mg, 0.12 mmol), triphosgene (22 mg, 74 µmol), triethylamine (45 µL, 0.33 mmol) and 3-(4-methylpiperazin-1-yl)propan-1-amine (41 µL, 0.24 mol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (51 mg, 43%).

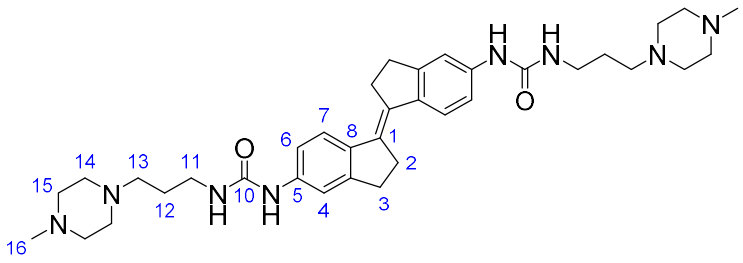
^{1}H NMR (500 MHz, D_2O) δ 7.62 (2H, d, *J* = 1.9 Hz, 7-C*H*), 7.33 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.09 (2H, dd, *J* = 8.1, 1.9 Hz, 5-C*H*), 3.69 (16H, br, 14-CH_2 and 15-CH_2), 3.37 – 3.34 (4H, m, 13-CH_2), 3.33 (4H, t, *J* = 5.8 Hz, 11-CH_2) 3.13 – 3.05 (8H, br, 2-CH_2 and 3-CH_2), 3.04 (6H, s, 16-CH_3), 2.05 – 1.98 (4H, m, 12-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 163.0 (q, *J* = 35.4 Hz, CF_3*C*O), 158.5 (10-*C*O), 143.4 (8-*C*), 143.0 (9-*C*), 136.4 (6-*C*), 135.7 (1-*C*), 125.3 (4-*C*H), 120.7 (5-*C*H), 117.4 (7-*C*H), 116.3 (q, *J* = 291.7 Hz, CF_3*C*O), 54.5 (13-*C*H_2), 50.2 (15-*C*H_2), 48.7 (14-*C*H_2), 42.8 (16-*C*H_3), 36.4 (11-*C*H_2), 31.6 (2-*C*H_2), 29.9 (3-*C*H_2), 24.4 (12-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3023 (w), 2931 (w), 2887 (w), 2842 (w), 1668 (s), 1647 (s), 1556 (m), 1483 (m), 1434 (m), 1416 (m), 1281 (m), 1241 (m), 1171 (s), 1119 (s), 1017 (m), 961 (m), 829 (s), 797 (s), 720 (s).

MALDI-HRMS $C_{36}H_{52}N_8O_2^+$ $[M+H]^+$ calcd: 629.4286, found 629.4305.

(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diyl)bis(3-(3-(4-methylpiperazin-1-yl)propyl)urea), E



. 4 TFA

Prepared according to General Procedure B using bis-aniline *E*-**49** (30 mg, 0.12 mmol), triphosgene (23 mg, 78 µmol), triethylamine (45 µL, 0.33 mmol) and 3-(4-methylpiperazin-1-yl)propan-1-amine (40 µL, 0.24 mmol). After lyophilisation following HPLC, the title compound was obtained (as the trifluoroacetate salt) as a brown solid (40 mg, 30%).

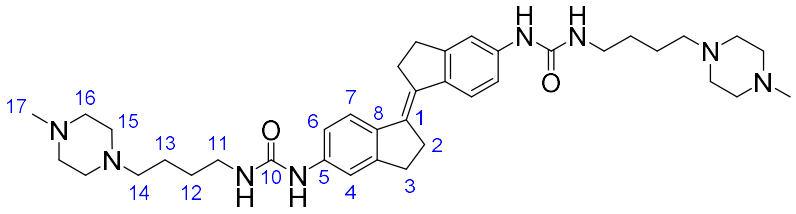
^{1}H NMR (500 MHz, D_2O) δ 7.54 (2H, d, *J* = 8.5 Hz, 7-C*H*), 7.29 (2H, d, *J* = 2.1 Hz, 4-C*H*), 7.16 (2H, dd, *J* = 8.5, 2.1 Hz, 6-C*H*), 3.57 – 3.33 (16H, br, 14-CH_2 and 15-CH_2), 3.29 (4H, t, *J* = 6.6 Hz, 11-CH_2), 3.12 (4H, t, *J* = 7.8 Hz, 13-CH_2), 3.05 (8H, br, 2-CH_2 and 3-CH_2), 2.92 (6H, s, 16-CH_3), 1.97 – 1.90 (4H, m, 12-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 163.0 (q, *J* = 35.3 Hz, CF_3*C*O), 158.0 (10-*C*O), 148.7 (9-*C*), 138.0 (8-*C*), 137.2 (5-*C*), 133.8 (1-*C*), 124.7 (7-*C*H), 118.8 (6-*C*H), 116.9 (4-*C*H), 116.3 (q, *J* = 291.8 Hz, *C*F_3CO), 54.4 (13-*C*H_2), 51.0 (15-*C*H_2), 49.1 (14-*C*H_2), 42.9 (16-*C*H_3), 36.7 (11-*C*H_2), 31.2 (2-*C*H_2), 30.4 (3-*C*H_2), 24.8 (11-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3308 (w), 3019 (w), 2928 (w), 1668 (s), 1636 (s), 1586 (s), 1565 (s), 1488 (m), 1428 (m), 1276 (m), 1238 (m), 1171 (s), 1123 (s), 1019 (m), 962 (m), 830 (s), 796 (s), 720 (s).

MALDI-HRMS $C_{36}H_{52}N_8O_2^+$ $[M+H]^+$ calcd: 629.4286, found 629.4299.

(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diyl)bis(3-(4-(4-methylpiperazin-1-yl)butyl)urea), F



. 4 TFA

Prepared according to General Procedure B using bis-aniline *E*-**49** (30 mg, 0.11 mmol), triphosgene (23 mg, 78 µmol), triethylamine (45 µL, 0.33 mmol) and 4-(4-methylpiperazin-1-yl)butan-1-amine (43 µL, 0.24 mmol). After lyophilisation following HPLC, the title compound was obtained (as the trifluoroacetate salt) as a brown solid (61 mg, 46%).

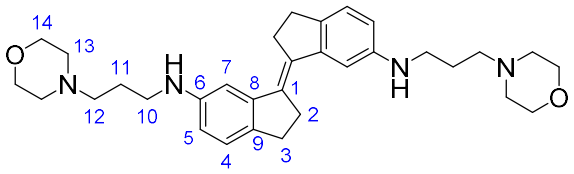
^{1}H NMR (500 MHz, D_2O) δ 7.43 (2H, d, *J* = 8.6 Hz, 7-C*H*), 7.23 (2H, d, *J* = 2.1 Hz, 4-C*H*), 7.10 (2H, dd, *J* = 8.6, 2.1 Hz, 6-C*H*), 3.42 – 3.25 (16H, br, 15-CH_2 and 16-CH_2), 3.19 (4H, t, *J* = 6.9 Hz, 11-CH_2), 3.04 – 2.99 (8H, m, 3-CH_2 and 14-CH_2), 2.92 (4H, br, 2-CH_2), 2.83 (6H, s, 17-CH_3), 1.74 – 1.67 (4H, m, 12-CH_2 or 13-CH_2), 1.58 (4H, app. p, *J* = 7.2 Hz, 12-CH_2 or 13-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 163.0 (q, *J* = 35.5 Hz, CF_3CO), 157.8 (10-*C*O), 148.5 (9-*C*), 137.8 (8-*C*), 137.3 (5-*C*), 133.6 (1-*C*), 124.6 (7-*C*H), 118.4 (6-*C*H), 116.4 (4-*C*H), 116.3 (q, *J* = 291.7 Hz, CF_3CO), 56.4 (14-*C*H_2), 51.1 (16-*C*H_2), 49.4 (15-*C*H_2), 43.0 (17-*C*H_3), 38.9 (11-*C*H_2), 31.1 (2-*C*H_2), 30.4 (3-*C*H_2), 26.6 (12-*C*H_2 or 13-*C*H_2), 21.4 (12-*C*H_2 or 13-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3321 (w), 3028 (w), 2933 (w), 2867 (w), 1670 (s), 1632 (s), 1581 (s), 1562 (s), 1487 (m), 1426 (m), 1268 (m), 1229 (m), 1175 (s), 1119 (s), 859 (m), 830 (s), 797 (s), 719 (s).

MALDI-HRMS $C_{38}H_{58}N_8O_2^+$ $[M+H]^+$ calcd: 657.4599, found 657.4579.

***(E)*-*N*^{6},*N*$^{6'}$-bis(3-morpholinopropyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, G**



. 4 TFA

Prepared according to General Procedure A using bromide *E*-**43** (51 mg, 0.13 mmol), BrettPhos Pd G3 (7.6 mg, 8.4 µmol), sodium *tert*-butoxide (49 mg, 0.51 mmol) and 3-

morpholinopropan-1-amine (42 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as an off-white solid (53 mg, 42%).

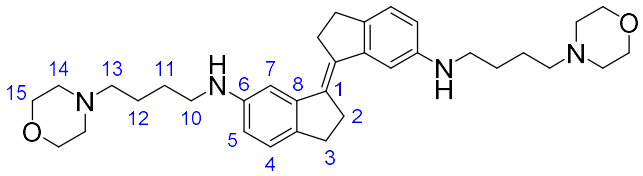
^{1}H NMR (500 MHz, D_2O) δ 7.64 (2H, d, *J* = 1.9 Hz, 7-C*H*), 7.55 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.31 (2H, dd, *J* = 8.1, 1.9 Hz, 5-C*H*), 4.19 – 4.07 (4H, br, 13-CH_2 or 14-CH_2), 3.90 – 3.78 (4H, br, 13-CH_2 or 14-CH_2), 3.57 (4H, t, *J* = 7.8 Hz, 10-CH_2), 3.57 – 3.47 (4H, br, 13-CH_2 or 14-CH_2), 3.30 (4H, t, *J* = 8.4 Hz, 12-CH_2), 3.24 – 3.18 (12H, br, 2-CH_2, 3-CH_2 and 13-CH_2 or 14-CH_2), 2.27 – 2.20 (4H, m, 11-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.7 Hz, CF_3CO), 148.7 (9-*C*), 144.1 (8-*C*), 136.3 (1-*C*), 134.1 (6-*C*), 126.7 (4-*C*H), 120.7 (5-*C*H), 117.5 (7-*C*H), 116.3 (q, *J* = 291.8 Hz, *C*F_3CO), 63.7 (14-*C*H_2), 53.8 (12-*C*H_2), 51.8 (13-*C*H_2), 47.6 (10-*C*H_2), 31.5 (2-*C*H_2), 30.2 (3-*C*H_2), 20.2 (11-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3017 (w), 2928 (w), 2869 (w), 2736 (w), 2641 (w), 2455 (w), 1664 (s), 1481 (m), 1461 (m), 1421 (m), 1195 (s), 1175 (s), 1117 (s), 831 (s), 797 (s), 720 (s).

ESI-HRMS for $C_{34}H_{48}N_4O_2$ $[M+H]^+$ calcd: 544.3772, found 544.3785.

(*E*)-*N*^{6},*N*$^{6'}$-bis(4-morpholinobutyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diamine, H



. 4 TFA

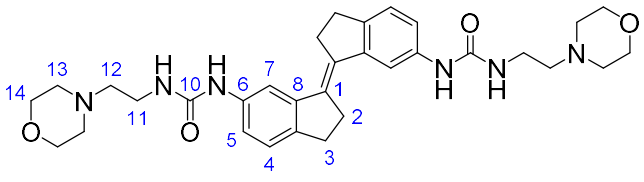
Prepared according to General Procedure A using bromide *E*-**43** (49 mg, 0.13 mmol), BrettPhos Pd G3 (6.4 mg, 7.1 µmol), sodium *tert*-butoxide (49 mg, 0.51 mmol) and 4-morpholinobutan-1-amine (47 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as an off-white solid (69 mg, 55%).

^{1}H NMR (500 MHz, D_2O) δ 7.67 (2H, d, *J* = 2.1 Hz, 7-C*H*), 7.57 (2H, d, *J* = 8.1 Hz, 4-C*H*), 7.34 (2H, dd, *J* = 8.1, 2.1 Hz, 5-C*H*), 4.14 (4H, d, *J* = 13.2 Hz, 14-CH_2 or 15-CH_2), 3.80 (4H, t, *J* = 12.8 Hz, 14-CH_2 or 15-CH_2), 3.57 – 3.49 (8H, m, 10-CH_2 and 14-CH_2 or 15-CH_2), 3.26 – 3.15 (8H, m, 13-CH_2 and 14-CH_2 or 15-CH_2), 3.22 (8H, app. s, 2-CH_2 and 3-CH_2), 1.90 – 1.80 (8H, m, 11-CH_2 and 12-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.5 Hz, CF_3CO), 149.3 (9-*C*), 144.0 (8-*C*), 136.3 (1-*C*), 133.4 (6-*C*), 126.7 (4-*C*H), 121.2 (5-*C*H), 118.1 (7-*C*H), 116.3 (q, *J* = 291.8 Hz, *C*F_3CO), 63.7 (15-*C*H_2), 56.2 (13-*C*H_2), 51.6 (14-*C*H_2), 50.7 (10-*C*H_2), 31.5 (2-*C*H_2), 30.2 (3-*C*H_2), 22.3 (11-*C*H_2 or 12-*C*H_2), 20.3 (11-*C*H_2 or 12-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 2928 (w), 2715 (w), 2638 (w), 2526 (w), 1671 (s), 1485 (m), 1455 (m), 1415 (m), 1198 (s), 1161 (s), 1128 (s), 1110 (s), 973 (m), 831 (s), 797 (s), 720 (s).
MALDI-HRMS $C_{32}H_{45}N_4O_2^+$ $[M]^+$ calcd: 517.3537, found 517.3543.

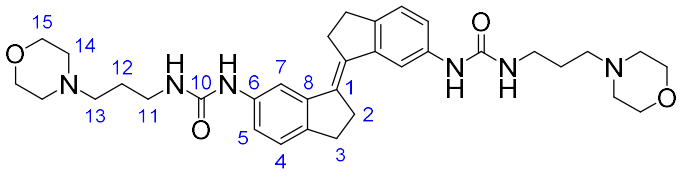
(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(3-(2-morpholinoethyl)urea), I



. 2 TFA

Prepared according to General Procedure B using bis-aniline *E*-**48** (29 mg, 0.11 mmol), triphosgene (22 mg, 74 µmol), triethylamine (45 µL, 0.33 mmol) and 2-morpholinoethan-1-amine (32 µL, 0.24 mmol). After lyophilisation following HPLC, the title compound was obtained (as the trifluoroacetate salt) as a brown solid (34 mg, 39%).
^{1}H NMR (500 MHz, D_2O) δ 7.51 (2H, d, *J* = 2.0 Hz, 7-C*H*), 7.23 (2H, d, *J* = 8.0 Hz, 4-C*H*), 7.00 (2H, dd, *J* = 8.0, 2.0 Hz, 5-C*H*), 4.14 (4H, app. d, *J* = 13.4 Hz, 13-CH_2 or 14-CH_2), 3.83 (4H, app. t, *J* = 12.6 Hz, 13-CH_2 or 14-CH_2), 3.65 – 3.62 (4H, br, 13-CH_2 or 14-CH_2), 3.61 (4H, t, *J* = 6.0 Hz, 11-CH_2), 3.35 (4H, t, *J* = 6.0 Hz, 12-CH_2), 3.25 (4H, app. t, *J* = 11.7 Hz, 13-CH_2 or 14-CH_2), 3.07 – 2.81 (8H, br, 2-CH_2 and 3-CH_2).
^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.4 Hz, CF_3*C*O), 158.2 (10-*C*O), 143.03 (8-*C*), 143.02 (9-*C*), 136.2 (6-*C*), 135.6 (1-*C*), 125.0 (4-*C*H), 119.8 (5-*C*H), 116.3 (q, *J* = 291.8 Hz, CF_3*C*O), 116.9 (7-*C*H), 63.8 (14-CH_2), 57.5 (12-*C*H_2), 52.0 (13-*C*H_2), 34.3 (11-*C*H_2), 31.5 (2-*C*H_2), 29.8 (3-*C*H_2).
ν_{max} / cm^{-1} (compressed solid) 3274 (w), 3056 (w), 2931 (w), 2878 (w), 1667 (s), 1590 (m), 1552 (m), 1484 (m), 1422 (m), 1199 (s), 1176(s), 1125 (s), 1049 (m), 1017 (m(, 832 (s), 799 (s), 721 (s).
ESI-HRMS $C_{32}H_{43}N_6O_4^+$ $[M+H]^+$ calcd: 575.3340, found 575.3342.

(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(3-(3-morpholinopropyl)urea), J



. 2 TFA

Prepared according to General Procedure B using bis-aniline *E*-**48** (29 mg, 0.11 mmol), triphosgene (22 mg, 74 µmol), triethylamine (45 µL, 0.33 mmol) and 3-morpholinopropan-1-amine (35 µL, 0.24 mmol). After lyophilisation following HPLC, the title compound was obtained as a brown solid (38 mg, 42%).

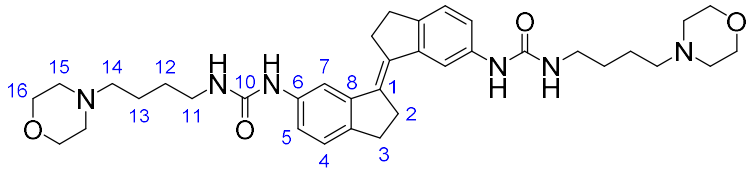
^{1}H NMR (500 MHz, D_2O) δ 7.49 (2H, s, 7-C*H*), 7.21 (2H, d, *J* = 8.0, 4-C*H*), 6.97 (2H, d, *J* = 8.0 Hz, 5-C*H*), 4.13 (4H, dd, *J* = 13.5, 3.5 Hz, 14-CH_2 or 15-CH_2), 3.81 (4H, ddd, *J* = 13.5, 11.9, 2.1 Hz, 14-CH_2 or 15-CH_2), 3.43 (4H, d, *J* = 12.7 Hz, 14-CH_2 or 15-CH_2), 3.30 (4H, t, *J* = 6.5 Hz, 11-CH_2), 3.24 – 3.20 (4H, m, 13-CH_2), 3.17 (4H, td, *J* = 12.7, 3.7 Hz,14-CH_2 or 15-CH_2), 2.93 – 2.83 (8H, br, 2-CH_2 and 3-CH_2), 2.08 – 1.84 (4H, m, 12-CH_2)

^{13}C NMR (126 MHz, D_2O): 162.9 (q, *J* = 35.6 Hz, CF_3CO), 158.2 (10-*C*), 143.0 (8-*C*), 142.8 (9-*C*), 136.4 (6-*C*), 135.5 (1-*C*), 125.1 (4-*C*H), 119.7 (5-*C*H), 116.7 (7-*C*H), 116.3 (q, *J* = 291.7 Hz, CF_3CO), 63.8 (15-*C*H_2), 54.7 (13-*C*H_2), 51.6 (14-*C*H_2), 36.4 (11-*C*H_2), 31.5 (2-*C*H_2), 29.8 (3-*C*H_2), 24.0 (12-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3313 (w), 3009 (w), 2931 (w), 2866 (w), 2632 (w), 2492 (w), 1664 (s), 1591 (m), 1546 (m), 1482 (m), 1420 (m), 1197 (s), 1124 (s), 907 (m), 875 (m), 831 (s), 798 (s), 720 (s).

MALDI-HRMS $C_{34}H_{46}N_6O_4^+$ $[M+H]^+$ calcd: 603.3653, found 603.3670.

(*E*)-1,1'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(3-(4-morpholinobutyl)urea), K



. 2 TFA

Prepared according to General Procedure B using bis-aniline *E*-**48** (30 mg, 0.11 mmol), triphosgene (23 mg, 78 µmol), triethylamine (45 µL, 0.33 mmol) and 4-morpholinobutan-1-

amine (39 µL, 0.24 mmol). After lyophilisation following HPLC, the title compound was obtained as a brown solid (33 mg, 35%).

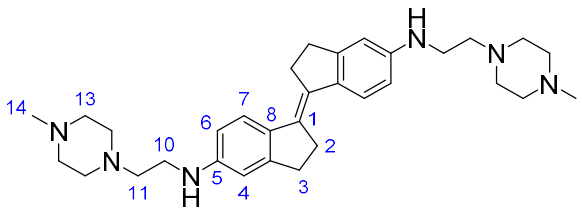
^{1}H NMR (500 MHz, D_2O) δ 7.39 (2H, d, *J* = 1.7 Hz, 7-C*H*), 7.04 (2H, d, *J* = 8.2 Hz, 4-C*H*), 6.92 (2H, dd, *J* = 8.2, 1.7 Hz, 5-C*H*), 4.02 (4H, dd, *J* = 13.1, 3.4 Hz, 15-CH_2 or 16-CH_2), 3.74 (4H, ddd, *J* = 13.7, 11.9, 2.1 Hz, 15-CH_2 or 16-CH_2), 3.56 – 3.32 (4H, m, 15-CH_2 or 16-CH_2), 3.14 (4H, t, *J* = 7.0 Hz, 11-CH_2), 3.12 – 2.84 (8H, m, 15-CH_2 or 16-CH_2 and 14-CH_2), 2.77 – 2.66 (8H, br, 2-CH_2 and 3-CH_2), 1.86 – 1.59 (4H, m, 12-CH_2 or 13-CH_2), 1.48 (4H, app. p, *J* = 7.2 Hz, 12-CH_2 or 13-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.8 (q, *J* = 35.5 Hz, CF_3CO), 157.8 (10-*C*O), 143.1 (8-*C*), 142.2 (9-*C*), 136.8 (6-*C*), 135.4 (1-*C*), 124.9 (4-*C*H), 119.3 (5-*C*H), 116.3 (q, *J* = 291.9 Hz, CF_3CO), 116.2 (7-*C*H), 63.7 (16-*C*H_2), 56.7 (14-*C*H_2), 51.4 (15-*C*H_2), 38.8 (11-*C*H_2), 31.4 (2-*C*H_2), 29.8 (3-*C*H_2), 26.4 (12-*C*H_2 or 13-*C*H_2), 20.5 (12-*C*H_2 or 13-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3310 (w), 3013 (w), 2928 (w), 2972 (w), 2644 (w), 2491 (w), 1667 (s), 1634 (s), 1587 (s), 1554 (s), 1482 (m), 1422 (m), 1449 (m), 1196 (s), 1176 (s), 1126 (s), 1053 (m), 1019 (m), 833 (s), 799 (s), 720 (s).

ESI-HRMS $C_{36}H_{51}N_6O_4^+$ $[M+H]^+$ calcd: 631.3966, found 631.3996.

(*E*)-*N*^{5},*N*$^{5'}$-bis(2-(4-methylpiperazin-1-yl)ethyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diamine, L



. 6 TFA

Prepared according to General Procedure A using bromide *E*-**45** (51 mg, 0.13 mmol), BrettPhos Pd G3 (7.3 mg, 8.1 µmol), sodium *tert*-butoxide (53 mg, 0.55 mmol) and 2-(4-methylpiperazin-1-yl)ethan-1-amine (44 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (75 mg, 48%).

^{1}H NMR (400 MHz, D_2O) δ 7.71 (2H, d, *J* = 8.5 Hz, 7-C*H*), 7.28 – 7.16 (2H, s, 4-C*H*), 7.10 (2H, d, *J* = 8.5 Hz, 6-C*H*), 3.78 – 3.21 (16H, br, 12-CH_2 and 13-CH_2), 3.65 (4H, t, *J* = 6.1 Hz, 10-CH_2), 3.16 (8H, app. s, 2-CH_2 and 3-CH_2), 3.12 (4H, t, *J* = 6.1 Hz, 11-CH_2), 2.97 (6H, s, 14-CH_3).

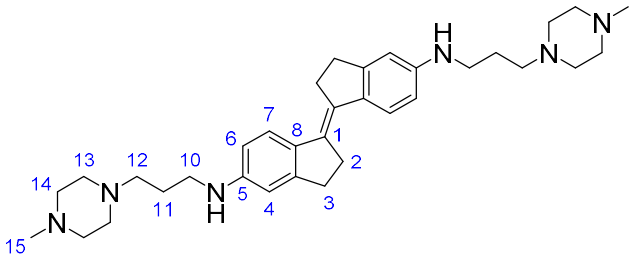
^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 36.0 Hz, CF_3CO), 149.9 (9-*C*), 139.3 (8-*C*), 138.6 (5-*C*), 134.3 (1-*C*), 125.5 (7-*C*H), 117.3 (6-*C*H), 116.3 (q, *J* = 291.6 Hz, CF_3CO), 115.1 (4-*C*H),

52.8 (11-*C*H), 51.9 (13-*C*H_2), 49.0 (12-*C*H_2), 43.5 (10-*C*H_2), 42.7 (14-*C*H_3), 31.4 (2-*C*H_2), 30.4 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3559 (w), 3013 (w), 2924 (w), 2846 (w), 1665 (s), 1608 (m), 1461 (m), 1411 (m), 1168 (s), 1120 (s), 1021 (m), 966 (m), 830 (s), 795 (s), 720 (s).

MALDI-HRMS $C_{32}H_{46}N_6^+$ $[M]^+$ calcd: 514.3778, found 514.3761.

(*E*)-*N*^{5},*N*$^{5'}$-bis(3-(4-methylpiperazin-1-yl)propyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diamine, M



. 6 TFA

Prepared according to General Procedure A using bromide *E*-**45** (49 mg, 0.13 mmol), BrettPhos Pd G3 (7.2 mg, 7.9 µmol), sodium *tert*-butoxide (49 mg, 0.51 mmol) and 3-(4-methylpiperazin-1-yl)propan-1-amine (49 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (73 mg, 47%).

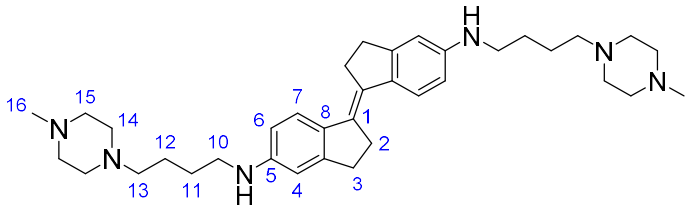
^{1}H NMR (400 MHz, D_2O) δ 7.63 (2H, d, *J* = 8.5 Hz, 7-C*H*), 7.19 (2H, s, 4-C*H*), 7.09 (2H, br, 6-C*H*), 3.60 – 3.32 (16H, m, 13-CH_2 and 14-CH_2), 3.36 (4H, t, *J* = 8.3 Hz, 10-CH_2), 3.19 – 3.11 (4H, br, 12-CH_2), 3.04 (8H, app. s, 2-CH_2 and 3-CH_2), 2.87 (6H, s, 15-CH_3), 2.09 – 1.98 (4H, m, 11-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.5 Hz, CF_3*C*O), 150.2 (9-*C*), 141.8 (8-*C*), 135.9 (5-*C*), 135.3 (1-*C*), 125.8 (7-*C*H), 119.0 (6-*C*H), 117.2 (4-*C*H), 116.3 (q, *J* = 291.7 Hz, *C*F_3CO), 53.7 (12-*C*H_2), 50.5 (14-*C*H_2), 48.9 (13*C*H_2), 46.6 (10-*C*H_2), 42.7 (15-*C*H_3), 31.4 (2-*C*H_2), 30.4 (3-*C*H_2), 21.1 (11-*C*H_2)

ν_{max} / cm^{-1} (compressed solid) 3024 (w), 2939 (w) 2846 (w), 1662 (s), 1489 (m), 1417 (m), 1278 (m), 1174 (s), 1116 (s), 963 (s), 830 (s), 796 (s), 719 (s).

MALDI-HRMS $C_{34}H_{50}N_6^+$ $[M]^+$ calcd: 542.4091, found 542.4082.

(*E*)-*N*^{5},*N*$^{5'}$-bis(4-(4-methylpiperazin-1-yl)butyl)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diamine, N



. 6 TFA

Prepared according to General Procedure A using bromide (*E*)-**11** (51 mg, 0.13 mmol), BrettPhos Pd G3 (7.2 mg, 7.9 µmol), sodium *tert*-butoxide (58 mg, 0.60 mmol) and 4-(4-methylpiperazin-1-yl)butan-1-amine (53 µL, 0.29 mmol). After lyophilisation following HPLC, the title compound (as the trifluoroacetate salt) was obtained as a brown solid (79 mg, 48%).

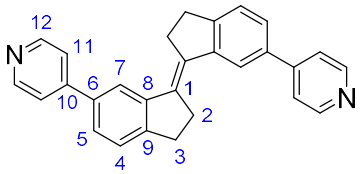
^{1}H NMR (500 MHz, D_2O) δ 7.81 (2H, d, *J* = 8.5 Hz, 7-C*H*), 7.41 (2H, s, 4-C*H*), 7.31 (2H, d, *J* = 8.5 Hz, 6-C*H*), 3.81 – 3.52 (16H, br, 14-CH_2 and 15-CH_2), 3.48 (4H, t, *J* = 7.5 Hz, 10-CH_2), 3.28 (4H, t, *J* = 7.8 Hz, 13-CH_2), 3.24 – 3.18 (8H, m, 2-CH_2 and 3-CH_2), 3.02 (6H, s, 16-CH_3), 1.95 – 1.78 (8H, m, 11-CH_2 and 12-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 162.9 (q, *J* = 35.5 Hz, CF_3CO), 150.2 (9-*C*), 142.8 (8-*C*), 135.6 (5-*C*), 134.6 (1-*C*), 125.8 (7-*C*H), 119.9 (6-*C*H), 118.3 (4-*C*H), 116.3 (q, *J* = 291.8 Hz, CF_3CO_2), 56.0 (12-CH_2), 50.3 (15-CH_2), 50.0 (10-CH_2), 48.7 (14-CH_2), 42.8 (16-CH_3), 31.4 (2-CH_2), 30.4 (3-CH_2), 22.5 (11-CH_2 or 12-CH_2), 20.8 (11-CH_2 or 12-CH_2).

ν_{max} / cm^{-1} (compressed solid) 3407 (w), 3015 (w), 2447 (m), 1667 (s), 1458 (m), 1415 (m), 1171 (s), 1120 (s), 1018 (m), 962 (m), 830 (s), 796 (s), 720 (s).

MALDI-HRMS $C_{36}H_{54}N_6^+$ $[M]^+$ calcd: 570.4404, found 570.4418.

(*E*)-6,6'-di(pyridin-4-yl)-2,2',3,3'-tetrahydro-1,1'-biindenylidene, *E*-58



Bromide *E*-**43** (245 mg, 0.628 mmol) and 4-pyridinylboronic acid (232 mg, 1.88 mmol) were suspended in toluene (22.5 mL) and EtOH (2.5 mL) and 2M aq. Na_2CO_3 (5 mL) was added. The solution was degassed by bubbling nitrogen for 30 min. $Pd(PPh_3)_4$ (144 mg, 0.123 mmol) was added in one portion and the mixture degassed for a further 10 min. The reaction was heated to 100 °C and stirred for 16 h. The mixture was then cooled to r.t., diluted with DCM (150 mL) and washed with H_2O (100 mL). The aqueous was back extracted with DCM (50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 0 – 5% MeOH in DCM, to afford a white solid. This was triturated in Et_2O (30 mL) and the remaining insoluble solid filtered and air-dried to afford the title compound as an off-white powder (128 mg, 53%).

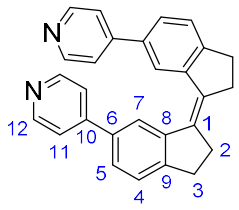
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.68 (4H, d, *J* = 6.2 Hz, 12-C*H*), 7.86 (2H, s, 7-C*H*), 7.56 (4H, d, *J* = 6.2 Hz, 11-C*H*), 7.50 (2H, d, *J* = 7.8 Hz, 5-C*H*), 7.46 (2H, d, *J* = 7.8 Hz, 4-C*H*), 3.34 – 3.29 (4H, br, 2-CH_2), 3.24 – 3.20 (4H, br, 3-CH_2).

^{13}C NMR (126 MHz, $CDCl_3$) δ 150.4 (12-*C*H), 149.1 (10-*C*), 148.5 (9-*C*), 144.2 (8-*C*), 136.9 (6-*C*), 135.9 (1-*C*), 126.3 (5-*C*H), 125.8 (4-*C*H), 123.2 (7-*C*H), 121.9 (11-*C*H), 32.4 (2-*C*H_2), 31.1 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3025 (w) 2942 (w), 2929 (w), 2911 (w), 2845 (w), 1593 (s), 1542 (m), 1474 (s), 1453 (w) 1434 (w), 1422 (w), 1440 (m), 1270 (m), 1069 (m) 992 (m), 887 (w), 839 (m), 812 (s), 800 (s), 729 (m), 607 (m), 535 (s).

ESI-HRMS for $C_{28}H_{23}N_2^+$ $[M+H]^+$ calcd: 387.1856, found: 387.1871.

(*Z*)-6,6'-di(pyridin-4-yl)-2,2',3,3'-tetrahydro-1,1'-biindenylidene, *Z*-58



Bromide *Z*-**43** (501 mg, 1.28 mmol) and 4-pyridinylboronic acid (495 mg, 4.02 mmol) were dissolved in toluene (22.5 mL) and EtOH (2.5 mL) and 2M aq. Na_2CO_3 (5 mL) was added. The solution was degassed by bubbling a stream of nitrogen for 30 min. $Pd(PPh_3)_4$ (147 mg, 0.127 mmol) was added in one portion and the mixture degassed for a further 10 min. The reaction was heated to 100 °C and stirred for 16 h. The mixture was then cooled to r.t., diluted with DCM (150 mL) and washed with H_2O (100 mL). The aqueous was back extracted with DCM (50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 0 – 5% MeOH in DCM, to afford a white solid. This was triturated in Et_2O (30 mL) and the remaining insoluble solid filtered and air-dried to afford the title compound as an off-white powder (257 mg, 52%).

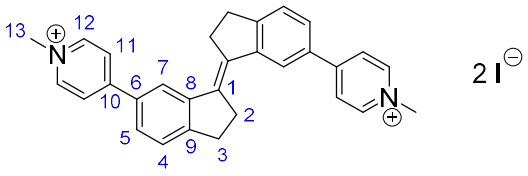
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.45 (4H, dd, *J* = 4.5, 1.8 Hz, 12-C*H*), 8.39 (2H, app. s, 7-C*H*), 7.44 (2H, dd, *J* = 7.8, 1.6 Hz, 5-C*H*), 7.41 (2H, d, *J* = 7.8 Hz, 4-C*H*), 7.29 (4H, dd, 4.5, 1.8 Hz, 11-C*H*), 3.09 – 3.05 (4H, m, 3-CH_2), 2.92 – 2.89 (4H, m, 2-CH_2).

^{13}C NMR (126 MHz, $CDCl_3$) δ 150.3 (12-*C*H), 149.6 (9-*C*), 148.7 (10-*C*), 141.5 (8-*C*), 136.0 (6-*C*), 135.4 (1-*C*), 126.4 (5-*C*H), 126.2 (4-*C*H), 122.1 (7-*C*H), 121.6 (11-*C*H), 34.9 (2-*C*H), 30.6 (3-*C*H).

ν_{max} / cm^{-1} (compressed solid) 3066 (w), 3027 (w), 2939 (w), 2905 (w), 2840 (w), 1596 (s), 1543 (m), 1472 (m), 1446 (w), 1422 (w), 1406 (w), 1268 (w), 1220 (w), 992 (w), 809 (s), 730 (m), 613 (w), 606 (w), 533 (w).

ESI-HRMS for $C_{28}H_{23}N_2^+$ $[M+H]^+$ calcd: 387.1856, found 387.1841.

(*E*)-4,4'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(1-methylpyridin-1-ium) iodide, (*E*)-56



Pyridine *E*-**58** (20 mg, 0.052 mmol), was dissolved in anhydrous DMF (2 mL) and methyl iodide (60 µL, 0.96 mmol) was added. The reaction was heated to 90 °C and stirred for 16 h. The mixture was then cooled to r.t. and concentrated *in vacuo*. The residue was washed with acetone (5 mL), filtered and the dark solid dried to afford the title compound (14 mg, 41%).

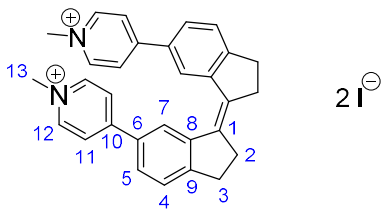
^{1}H NMR (500 MHz, DMSO-d_6) δ 9.00 (4H, d, *J* = 6.9 Hz, 12-C*H*), 8.56 (4H, d, *J* = 6.9 Hz, 11-C*H*), 8.11 (2H, d, *J* = 1.6 Hz, 7-C*H*), 7.95 (2H, dd, *J* = 7.9, 1.6 Hz, 5-C*H*), 7.66 (2H, d, *J* = 7.9 Hz, 4-C*H*), 4.36 (6H, s, 13-CH_3), 3.40 – 3.36 (4H, m, 2-CH_2), 3.25 – 3.22 (4H, m, 3-CH_2).

^{13}C NMR (126 MHz, DMSO-d_6) δ 154.8 (10-*C*), 151.6 (9-*C*), 145.5 (12-*C*H), 143.8 (8-*C*), 135.7 (1-*C*), 132.4 (6-*C*H_2), 127.6 (5-*C*H), 126.4 (4-*C*H), 124.3 (11-*C*H), 123.6 (7-*C*H), 47.1 (13-*C*H_3), 31.4 (2-*C*H_2), 30.7 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3121 (w), 3088 (w), 3038 (w), 2930 (w), 2910 (w), 2844 (w), 1636 (s), 1601 (s), 1558 (m), 1522 (m), 1472 (s), 1439 (m), 1414 (m), 1348 (w), 1312 (m) 1298 (m), 1289 (m), 1272 (m), 1223 (w), 1198 (s), 856 (w), 820 (s), 800 (m), 723 (m), 607 (m), 587 (m), 534 (m), 510 (s), 426 (s).

MALDI-HRMS $C_{29}H_{25}N_2^+$ [M-CH_3]$^+$ calcd: 401.2012, found 401.2019.

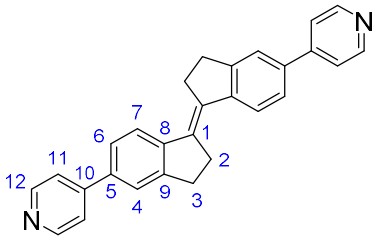
(*Z*)-4,4'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diyl)bis(1-methylpyridin-1-ium) iodide, *Z*-56



Pyridine *Z*-**58** (20 mg, 0.052 mmol), was dissolved in anhydrous DMF (2 mL) and methyl iodide (60 µL, 0.96 mmol) was added. The reaction was heated to 90 °C and stirred for 1 h. The mixture was then cooled to r.t. and concentrated *in vacuo*. The residue was washed with acetone (5 mL), filtered and the dark solid dried to afford the title compound (22 mg, 65%)

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.82 (4H, d, *J* = 6.5 Hz, 12-C*H*), 8.64 (2H, d, *J* = 1.9 Hz, 7-C*H*), 8.31 (4H, d, *J* = 6.5 Hz, 11-C*H*), 7.97 (2H, dd, *J* = 7.9, 1.9 Hz, 5-C*H*), 7.66 (2H, d, *J* = 7.9 Hz, 4-C*H*), 4.28 (6H, s, 13-CH_3), 3.13 – 3.08 (4H, m, 3-CH_2), 2.95 – 2.92 (4H, m, 2-CH_2).
^{13}C NMR (126 MHz, DMSO-d_6) δ 154.3 (10-*C*), 153.0 (9-*C*), 145.4 (12-*C*H), 141.1 (8-*C*), 135.5 (1-*C*), 131.4 (6-*C*), 127.8 (5-*C*H), 126.9 (4-*C*H), 123.5 (11-*C*H), 121.9 (7-*C*H), 47.0 (13-*C*H_3), 34.5 (2-*C*H_2), 30.1 (3-*C*H_2).
ν_{max} / cm^{-1} (compressed solid) 3015 (w), 2929 (w), 1639 (s), 1594 (m), 1558 (m), 1522 (m), 1473 (m), 1443 (m), 1425 (m), 1321 (m), 1279 (m), 1203 (m), 1194 (m), 819 (S), 722 (w), 599 (m), 508 (s), 425 (s).
ESI-HRMS $C_{30}H_{28}N_2^{2+}$ $[M]^{2+}$ calcd: 208.1121, found 208.1125.

(*E*)-5,5'-di(pyridin-4-yl)-2,2',3,3'-tetrahydro-1,1'-biindenylidene, *E*-59

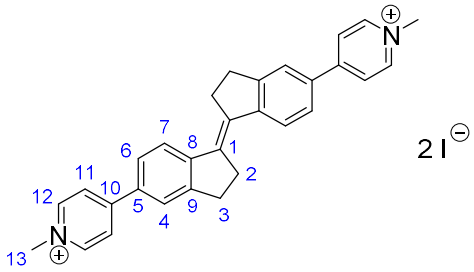


Bromide *E*-**45** (250 mg, 0.641 mmol) and 4-pyridinylboronic acid (236 mg, 1.92 mmol) were dissolved in toluene (23 mL) and EtOH (2.5 mL) and 2M aq. Na_2CO_3 (5 mL) were added. The solution was degassed by bubbling nitrogen for 30 min. $Pd(PPh_3)_4$ (74 mg, 0.064 mmol) was added in one portion and the mixture degassed for a further 10 min. The reaction was heated to 100 °C and stirred for 16 h. The mixture was then cooled to r.t., diluted with DCM (150 mL) and washed with H_2O (100 mL). The aqueous was back extracted with DCM (50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 0 – 5% MeOH in DCM, to afford the title compound a yellow solid (190 mg, 76%).
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.66 (4H, d, *J* = 6.2 Hz, 12-C*H*), 7.72 (2H, d, *J* = 8.2 Hz, 7-C*H*), 7.62 (2H, d, *J* = 1.7 Hz, 4-C*H*), 7.56 (2H, dd, *J* = 8.2, 1.7 Hz, 6-C*H*), 7.54 (4H, d, *J* = 6.2 Hz, 11-C*H*), 3.24 (8H, app. s, 2-CH_2 and 3-CH_2).
^{13}C NMR (101 MHz, $CDCl_3$) δ 150.4 (12-*C*H), 148.5 (9-*C*), 148.2 (10-*C*), 144.2 (8-*C*), 136.9 (5-*C*), 136.2 (1-*C*), 125.6 (6-*C*H), 125.2 (7-*C*H), 123.6 (4-*C*H), 121.6 (11-*C*H), 32.2 (2-*C*H_2), 31.2 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3485 (w), 3027 (w), 2920 (w), 2844 (w), 1594 (s), 1539 (m), 1476 (m), 1428 (w), 1407 (m), 1308 (w), 1298 (w), 1258 (w), 1221 (w), 997 (w), 888 (w), 809 (s), 731 (m), 719 (m), 613 (w), 527 (w).

ESI-HRMS for $C_{28}H_{23}N_2^+$ $[M+H]^+$ calcd: 387.1856, found: 387.1872.

(*E*)-4,4'-(2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-5,5'-diyl)bis(1-methylpyridin-1-ium) iodide



Pyridine *E*-**59** (50 mg, 0.13 mmol) was dissolved in anhydrous DMF (2 mL) and the solution heated to 90 °C. Methyl iodide (24 µL, 0.34 mmol) was added and the reaction stirred for 1 h. The mixture was then cooled to r.t. and concentrated *in vacuo.* The residue was suspended in 1:1 H_2O/MeCN (5 mL) and the precipitate filtered and dried under vacuum to afford the title compound as a bright orange powder (47 mg, 55%).

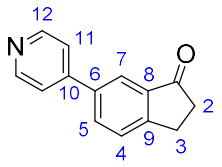
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.99 (4H, d, *J* = 6.4 Hz, 12-C*H*), 8.54 (4H, d, *J* = 6.4 Hz, 11-C*H*), 8.15 (2H, s, 4-C*H*), 8.05 (2H, d, *J* = 8.5 Hz, 6-C*H*), 7.87 (2H, d, *J* = 8.5 Hz, 7-C*H*), 4.32 (6H, s, 15-CH_3) 3.33 (4H, s, 2-CH_2), 3.27 (4H, s, 3-CH_2).

^{13}C NMR (126 MHz, DMSO-d_6) δ 154.1 (10-*C*), 149.4 (9-*C*), 146.4 (8-*C*), 145.9 (12-*C*H), 137.9 (1-*C*), 132.7 (5-*C*), 127.3 (6-*C*H), 125.9 (7-*C*H), 125.0 (4-*C*H), 124.1 (11-*C*H), 47.4 (13-CH_3), 32.1 (2-*C*H_2), 30.9 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3020 (w), 2921 (w), 2882 (w), 1637 (s), 1600 (s), 1574 (m), 1556 (m), 1525 (s), 1485 (s), 1473 (m), 1448 (m), 1422 (m), 1352 (m), 1316 (m), 1306 (m), 1285 (m), 1239 (m), 1191 (s), 1161 (m), 1122 (m), 1053 (m), 822 (m), 849 (m), 829 (s), 795 (s), 714 (m), 666 (m), 508 (s).

ESI-HRMS for $C_{29}H_{25}N_2^+$ $[M-CH_3]^+$ calcd: 401.2012, found 401.2018.

6-(pyridin-4-yl)-2,3-dihydro-1H-inden-1-one, 62



6-Bromo-1-indanone (249 mg, 1.18 mmol) and 4-pyridinylboronic acid (256 mg, 2.08 mmol) were dissolved in toluene (45 mL) and EtOH (5 mL) and 2M aq. Na_2CO_3 (10 mL) were added. The solution was degassed by bubbling nitrogen for 30 min. $Pd(PPh_3)_4$ (137 mg, 0.119 mmol) was added in one portion and the mixture degassed for a further 10 min. The reaction was heated to 100 °C and stirred for 17 h. The mixture was then cooled to r.t., diluted with DCM (50 mL) and washed with H_2O (50 mL). The aqueous was back extracted with DCM (2 x 50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 0 – 5% MeOH in DCM, to afford a white solid. This was triturated in Et_2O (20 mL) and the remaining insoluble solid filtered and air-dried to afford the title compound as an off-white powder (70 mg, 28%).

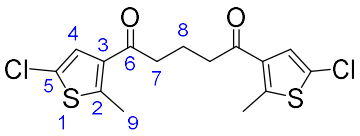
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.66 (2H, dd, *J* = 4.6, 1.6 Hz, 12-C*H*), 8.12 (1H, dd, *J* = 8.0 1.9 Hz, 5-C*H*), 8.01 (1H, d, *J* = 1.9 Hz, 7-C*H*), 7.79 (2H, dd, *J* = 4.6, 1.6 Hz, 11-C*H*), 7.75 (1H, d, *J* = 8.0 Hz, 4-C*H*), 3.19 – 3.16 (2H, m, 3-CH_2), 2.73 – 2.70 (2H, m, 2-CH_2).

^{13}C NMR (126 MHz, DMSO-d_6) δ 206.0 (1-*C*O), 156.2 (9-*C*), 150.3 (12-*C*H), 146.1 (10-*C*), 137.6 (8-*C*), 136.5 (6-*C*), 133.2 (5-C*H*), 127.9 (4-*C*H), 121.4 (11-*C*H), 120.9 (7-*C*H), 36.3 (2-*C*H_2), 25.4 (3-*C*H_2).

ν_{max} / cm^{-1} (compressed solid) 3058 (w), 3032 (w), 2956 (w), 2924 (w), 2915 (w), 1700 (s), 1615 (m), 1594 (s), 1548 (m), 1481 (m), 1437 (m), 1402 (m), 1302 (m), 1267 (m), 1237 (m), 1178 (m), 1118 (m), 1071 (m), 1043 (m), 1028 (m), 991 (m), 907 (w), 818 (s), 803 (s), 647 (m), 560 (s), 468 (s).

ESI-HRMS for $C_{14}H_{12}NO^+$ $[M+H]^+$ calcd: 210.0913, found: 210.0914.

1,5-bis(5-chloro-2-methylthiophen-3-yl)pentane-1,5-dione, 82



2-Chloro-5-methylthiophene (3.14 g, 23.6 mmol) and glutaryl chloride (2.00 g, 11.8 mmol) were dissolved in DCM (150 mL) and anhydrous aluminium chloride (4.57 g, 34.3 mmol) was added. The solution was stirred for at r.t. for 18 h, then poured onto ice water (200 mL). The layers were separated and the aqueous back extracted with DCM (2 x 100 mL). The combined organic extractions were dried ($MgSO_4$), filtered and concentrated *in vacuo* to afford the title compound as a pale brown solid (3.60 g, 92%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.18 (2H, s, 4-C*H*), 2.85 (4H, t, *J* = 7.0 Hz, 7-CH_2), 2.66 (6H, s, 9-CH_3), 2.05 (2H, p, *J* = 7.0 Hz, 8-CH_2).

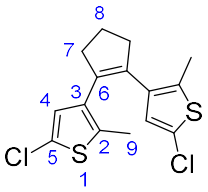
^{13}C NMR (101 MHz, $CDCl_3$) δ 195.0 (6-*C*O), 147.9 (Ar*C*), 135.0 (Ar*C*), 127.0 (4-*C*H), 125.4 (Ar*C*), 40.6 (7-*C*H_2), 18.3 (8-*C*H_2), 16.2 (9-*C*H_3).

ν_{max} / cm^{-1} (compressed solid) 3093 (w), 2961 (w), 2923 (w), 2896 (w), 1673 (s), 1524 (m), 1455 (s), 1409 (m), 1374 (m), 1243 (m), 1203 (s), 1156 (m), 1130 (m), 1006 (m), 970 (m), 822 (m), 784 (m), 729 (m), 483 (s).

ESI-LRMS for $C_{15}H_{14}Cl_2O_2S_2Na^+$ $[M+Na]^+$ calcd: 382.9, found: 383.0.

Proton and carbon NMR were consistent with literature data.[240]

1,2-bis(5-chloro-2-methylthiophen-3-yl)cyclopent-1-ene, 83



Zinc powder (1.09 g, 16.6 mmol) was suspended in anhydrous THF (150 mL). $TiCl_4$ (1.5 mL, 13 mmol) was added dropwise as the solution was stirred vigorously. The solution was heated to reflux and stirred for 1 h. The mixture was then cooled to 0 °C and thiophene **82** (2.00 g, 5.54 mmol) was added to the suspension. The mixture was heated to reflux and stirred for 4 h, then quenched with sat. aq. Na_2CO_3 (10 mL) causing the evolution of a white precipitate. The solution was filtered through Celite and washed with EtOAc (20 mL). The filtrate was dried

($MgSO_4$), concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with hexane, to afford the title compound as a pale yellow solid (1.20 g, 66%).

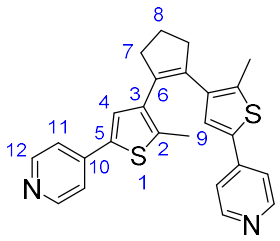
^{1}H NMR (400 MHz, $CDCl_3$) δ 6.58 (2H, s, 4-C*H*), 2.72 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.02 (2H, p, *J* = 7.5 Hz, 8-CH_2), 1.89 (6H, s, 9-CH_3).

^{13}C NMR (101 MHz, $CDCl_3$) δ 135.0 (Ar*C*), 134.6 (5-*C*), 133.4 (Ar*C*), 126.8 (4-*C*H), 125.3 (Ar*C*), 38.5 (7-*C*H_2), 23.0 (8-*C*H_2), 14.3 (9-*C*H_3).

ν_{max} / cm^{-1} (compressed solid) 2950 (m), 2919 (m), 2844 (m), 1553 (w), 1538 (w), 1457 (s), 1439 (s), 1378 (w), 1310 (w), 1291 (w), 1202 (m), 1159 (m), 1144 (m), 1055 (w), 1014 (m), 922 (s), 964 (w), 827 (s), 748 (w), 667 (w), 651 (w), 482 (s).

Proton and carbon NMR were consistent with literature data.[240]

1,2-bis(2-methyl-5-(pyridin-4-yl)thiophen-3-yl)cyclopent-1-ene, 85

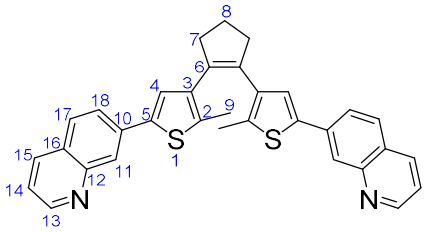


Thiophene **83** (245 mg, 0.744 mmol) was dissolved in anhydrous THF (5 mL) and the solution cooled to 0 °C. *n*-BuLi (2.5 M in hexanes, 655 µL, 1.64 mmol) was added dropwise, and the solution stirred for 30 min. *tris*-*n*-butylborate (455 µL, 2.23 mmol) was added and the resulting solution warmed to r.t. and stirred for 1 h to afford crude boronic ester **84**. Separately, $Pd(PPh_3)_4$ (86 mg, 0.074 mmol) and 4-bromopyridine hydrochloride (318 mg, 1.63 mmol) were dissolved in 2.5 M aq. K_2CO_3 (5 mL) and the solution was degassed by evacuation under vacuum and backfilling with nitrogen three times. The boronic ester solution was then added to this mixture and the resulting solution heated to 70 °C and stirred for 18 h. The mixture was cooled, diluted with DCM (50 mL) and washed with water (50 mL). The aqueous was back extracted with DCM (2 x 50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 2:1 hexane:EtOAc, to yield a pale violet solid. This was further purified by trituration in 1:1 Et_2O:pentane to afford the title compound as a purple amorphous solid (81 mg, 26%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.53 (4H, d, *J* = 5.2 Hz, 12-C*H*), 7.34 (4H, d, *J* = 5.2 Hz, 11-C*H*), 7.21 (2H, s, 4-C*H*), 2.85 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.11 (2H, p, *J* = 7.5 Hz, 8-CH_2), 2.02 (6H, s, 9-CH_3).

^{13}C NMR (101 MHz, $CDCl_3$) δ 150.4 (12-*C*H), 141.4 (10-*C*), 137.4 (Ar*C*), 137.2 (Ar*C*), 136.8 (Ar*C*), 135.0 (6-*C*), 126.4 (4-*C*H), 119.4 (11-*C*H), 38.6 (7-*C*H_2), 23.1 (8-*C*H_2), 14.8 (9-*C*H_3).
ν_{max} / cm^{-1} (compressed solid) 3065 (w), 3028 (w), 2952 (w), 2920 (w), 2946 (w), 1596 (s), 1540 (w), 1498 (m), 1461 (m), 1437 (m), 1411 (m), 1378 (w), 1310 (w), 1294 (w), 1220 (m), 1171 (w), 1120 (w), 992 (m), 909 (w), 857 (w), 814 (s), 729 (m), 601 (m), 542 (m), 4723 (m).
ESI-LRMS for $C_{25}H_{23}N_2S_2^+$ $[M+H]^+$ calcd: 415.1, found: 415.1.
Proton and carbon NMR were consistent with literature data.[224]

1,2-bis(2-methyl-5-(quinolin-7-yl)thiophen-3-yl)cyclopent-1-ene, 86

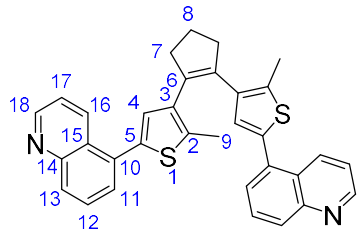


Synthesised by Dr Javier Ramos Soriano
Thiophene **83** (150 mg, 0.46 mmol) was dissolved in anhydrous THF (3 mL) and the solution cooled to 0 °C. *n*-BuLi (2.5 M in hexanes, 402 µL, 1.01 mmol) was added dropwise, and the solution stirred for 30 min. *tris*-*n*-butylborate (370 µL, 1.37 mmol) was added and the resulting solution warmed to r.t. and stirred for 1 h to afford crude boronic ester **84**. Separately, $Pd(PPh_3)_4$ (54 mg, 0.046 mmol) was suspended in dry THF (2.5 mL) and the resulting mixture was heated at 70 °C. After 30 min, aq. 2.5 M K_2CO_3 (3 mL), ethylene glycol (1 drop) and 7-bromoquinoline (220 mg, 1.01 mmol) were added. The boronic ester solution was then added to this mixture and the resulting solution heated to 70 °C and stirred for 19 h. The mixture was cooled and water (10 mL) was added. The aqueous layer was extracted with Et_2O (3 x 20 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 1:1 hexane:EtOAc, to afforded the title compound as a purple amorphous solid (188 mg, 80%).
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.88 (2H, dd, *J* = 4.2, 1.7 Hz, 13-C*H*), 8.21 (2H, m, 11-C*H*), 8.09 (2H, ddd, *J* = 8.2, 1.7, 0.8 Hz, 15-C*H*), 7.76 (2H, d, *J* = 8.3 Hz, 17-C*H*), 7.71 (2H, dd, *J* = 8.3, 1.8 Hz, 18-C*H*), 7.33 (2H, dd, *J* = 8.2, 4.2 Hz, 14-C*H*), 7.26 (2H, under $CDCl_3$, 4-C*H*), 2.89 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.14 (2H, p, *J* = 7.5 Hz, 8-CH_2), 2.07 (6H, s, 9-CH_3)
^{13}C NMR (101 MHz, $CDCl_3$) δ 151.0 (13-*C*H), 148.6 (12-*C*), 139.0 (5-*C*), 137.1 (2-*C*), 136.0 (3-*C*), 135.7 (15-*C*H), 135.5 (10-*C*), 134.8 (6-*C*), 128.2 (17-*C*H), 127.3 (16-*C*), 125.3 (4-*C*), 124.6 (18-*C*H), 124.5 (11-*C*H), 120.7 (14-*C*H), 38.6 (7-*C*H_2), 23.0 (8-*C*H_2), 14.6 (9-*C*H_3); ESI-

ν_{max} / cm^{-1} (thin film) 3053 (w), 3011 (w), 2922 (m), 2852 (w), 1617 (s), 1596 (m), 1502 (m), 1439 (m), 1392 (w), 1321 (w), 1176 (w), 1152 (w), 1123 (w), 1128 (w), 1092 (w), 940 (w), 887 (m), 830 (s), 767 (m), 736 (w), 665 (w), 644 (w)

ESI-HRMS for $C_{33}H_{27}N_2S_2^+$ $[M+H]^+$ calcd: 515.1610, found: 515.1602.

1,2-bis(2-methyl-5-(quinolin-5-yl)thiophen-3-yl)cyclopent-1-ene, 87



Thiophene **83** (245 mg, 0.744 mmol) was dissolved in anhydrous THF (5 mL) and the solution cooled to 0 °C. *n*-BuLi (2.5 M in hexanes, 655 µL, 1.64 mmol) was added dropwise, and the solution stirred for 30 min. *tris*-*n*-butylborate (455 µL, 2.23 mmol) was added and the resulting solution warmed to r.t. and stirred for 1 h to afford crude boronic ester **84**. Separately, $Pd(PPh_3)_4$ (86 mg, 0.074 mmol) and 5-bromoquinoline (341 mg, 1.63 mmol) were dissolved in 2.5 M aq. K_2CO_3 (5 mL) and the solution was degassed by evacuation under vacuum and backfilling with nitrogen three times. The boronic ester solution was then added to this mixture and the resulting solution heated to 70 °C and stirred for 18 h. The mixture was cooled and diluted with DCM (50 mL) and washed with water (50 mL). The aqueous was back extracted with DCM (2 x 50 mL) and the combined organic extractions dried ($MgSO_4$) and filtered. The filtrate was concentrated *in vacuo* and the residue purified by flash silica chromatography, eluting with 2:1 hexane:EtOAc, afford the title compound as a pale red oil (113 mg, 30%).

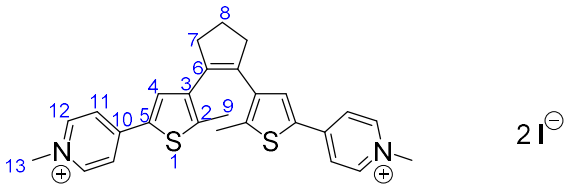
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.84 (2H, dd, *J* = 4.2, 1.7 Hz, 18-C*H*), 8.39 (2H, ddd, *J* = 8.7, 1.7, 0.9 Hz, 16-C*H*), 8.07 (2H, dt, *J* = 8.5, 1.1 Hz, 13-C*H*), 7.68 (2H, dd, *J* = 8.5, 7.2 Hz, 12-C*H*), 7.57 (2H, dd, *J* = 7.2, 1.1 Hz, 11-C*H*), 7.23 (2H, dd, *J* = 8.7, 4.2 Hz, 17-C*H*), 6.90 (2H, s, 4-C*H*), 2.89 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.26 (6H, s, 9-CH_3), 2.12 (2H, p, *J* = 7.5 Hz, 8-CH_2).

^{13}C NMR (101 MHz, $CDCl_3$) δ 150.5 (18-*C*H), 148.7 (14-*C*), 136.6 (2-*C*), 136.0 (5-*C*), 135.8 (3-*C*), 135.4 (6-*C*), 134.2 (16-*C*H), 132.9 (10-*C*), 129.4 (13-*C*H and 4-*C*H), 128.9 (12-*C*H), 128.0 (11-*C*H), 126.8 (15-*C*), 121.4 (17-*C*H), 38.2 (7-*C*H_2), 23.3 (8-*C*H_2), 14.5 (9-*C*H_3).

ν_{max} / cm^{-1} (thin film) 3060 (w), 3031 (w), 2949 (w), 2916 (w), 2845 (w), 1592 (m), 1570 (m), 1499 (m), 1484 (m), 1437 (m), 1393 (m), 1309 (m), 1197 (w), 1161 (w), 1146 (w), 1120 (w), 1092 (w), 930 (m), 908 (m), 854 (w), 828 (m), 802 (s), 730 (s), 694 (m), 667 (w), 644 (w), 585 (m), 542 (m).

ESI-HRMS for $C_{33}H_{27}N_2S_2^+$ $[M+H]^+$ calcd: 515.1610, found: 515.1626.

4,4'-(cyclopent-1-ene-1,2-diylbis(5-methylthiophene-4,2-diyl))bis(1-methylpyridin-1-ium) iodide, 77o



Pyridine **85** (40 mg, 0.096 mmol) was dissolved in DCM (1 mL). Methyl iodide (13 µL, 0.21 mmol) was added and the solution stirred at r.t. for 3 h. The solvent was removed *in vacuo*, the residue washed with 1:1 hexane:EtOAc and the precipitate filtered and dried under vacuum to afford the title compound as a green powder (52 mg, 78%).

^{1}H NMR (500 MHz, D_2O) δ 8.51 (4H, d, *J* = 6.7 Hz, 12-C*H*), 7.97 (4H, d, *J* = 6.7 Hz, 11-C*H*), 7.84 (2H, s, 4-C*H*), 4.24 (6H, s, 13-CH_3), 2.89 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.14 (2H, p, *J* = 7.5 Hz, 8-CH_2), 2.06 (6H, s, 9-CH_3).

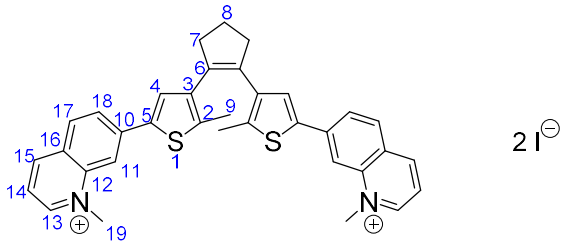
^{13}C NMR (126 MHz, D_2O) δ 148.7 (10-*C*), 144.9 (Ar*C*), 144.4 (12-*C*H), 138.8 (Ar*C*), 135.3 (3-*C*), 132.9 (5-*C*), 132.5 (Ar*C*), 121.5 (11-*C*H), 46.8 (13-*C*H_3), 37.9 (7-*C*H_2), 22.7 (8-*C*H_2), 14.1 (9-*C*H_3).

ν_{max} / cm^{-1} (compressed solid) 3026 (w), 2946 (w), 2922 (w), 2841 (w), 1634, (s), 1595 (m), 1533 (m), 1511 (s), 1471 (m), 1435 (m), 1416 (m), 1372 (w), 1342 (w), 1313 (w), 1293 (w), 1260 (w), 1220 (m), 1190 (s), 1181 (s), 1150 (m), 1069 (w), 1047 (w), 1028 (w), 993 (w), 920 (w), 893 (w), 830 (s), 756 (w), 720 (w), 541 (m), 471 (s).

ESI-LRMS for $C_{27}H_{28}N_2S_2^{2+}$ $[M]^{2+}$ calcd: 222.1, found: 222.1.

Proton NMR were consistent with literature data.[241]

7,7'-(cyclopent-1-ene-1,2-diylbis(5-methylthiophene-4,2-diyl))bis(1-methylquinolin-1-ium) iodide, 78o



Synthesised by Dr Javier Ramos Soriano

Quinoline **86** (100 mg, 0.19 mmol) was dissolved in in acetonitrile (5 mL) and methyl iodide (73 µL, 1.2 mmol) was added. The solution was stirred in a sealed vessel at 70 °C for 16 h. The solvent was removed *in vacuo* and the residue was washed with Et_2O and $CHCl_3$ to afford the title compound as a green powder (108 mg, 70%).

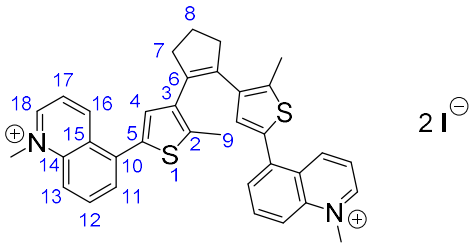
^{1}H NMR (400 MHz, DMSO-d_6) δ 9.41 (2H, d, *J* = 5.8 Hz, 13-C*H*), 9.17 (2H, d, *J* = 8.2 Hz, 15-C*H*), 8.47 – 8.41 (4H, m, 11-C*H*, 17-C*H*), 8.26 (2H, dd, *J* = 8.6, 1.6 Hz, 18-C*H*), 8.09 – 8.01 (4H, m, 4-C*H*, 14-C*H*), 4.63 (6H, s, 19-CH_3), 2.96 (4H, t, *J* = 7.5 Hz, 7-CH_2), 2.14 (2H, p, *J* = 7.5 Hz, 8-CH_2), 2.00 (6H, s, 9-CH_3).

^{13}C NMR (101 MHz, DMSO-d_6) δ 150.2 (13-*C*H), 146.2 (15-*C*H), 139.8 (10-*C*), 139.2 (12-*C*), 138.6 (2-*C*), 137.8 (3-*C*), 136.6 (5-*C*), 134.5 (6-*C*), 131.3 (17-*C*H), 129.5 (4-*C*), 128.2 (16-*C*), 127.2 (18-*C*H), 121.2 (14-*C*H), 112.5 (11-*C*H), 45.3 (19-*C*H_3), 38.3 (7-*C*H_2), 22.3 (8-*C*H_2), 14.4 (9-*C*H_3).

ν_{max} / cm^{-1} (compressed solid) 3054 (w), 3012 (w), 2944 (w), 1752, (m), 1626 (s), 1593 (m), 1512 (s), 1441 (m), 1370 (m), 1215 (m), 838 (s)

ESI-HRMS for $C_{35}H_{32}IN_2S_2^+$ $[M]^+$ calcd: 671.1046, found: 671.1033.

4,4'-(cyclopent-1-ene-1,2-diylbis(5-methylthiophene-4,2-diyl))bis(1-methylpyridin-1-ium) iodide, 79o



Quinoline **87** (50 mg, 0.097 mmol) was dissolved in DCM (1 mL) and methyl iodide (18 µL, 0.29 mmol) was added. The solution was stirred in a sealed vessel at 50 °C for 16 h. The solvent was removed *in vacuo*, the residue was washed with 1:1 hexane:EtOAc and the precipitate filtered and dried under vacuum to afford the title compound as an orange powder (64 mg, 83%).

^{1}H NMR (500 MHz, D_2O) δ 9.13 (2H, d, *J* = 5.8 Hz, 18-C*H*), 9.11 (2H, d, *J* = 8.7 Hz, 16-C*H*), 8.34 (2H, dt, *J* = 9.0, 1.0 Hz, 13-C*H*), 8.20 (2H, dd, *J* = 9.0, 7.4 Hz, 12-C*H*), 7.97 (2H, dd, *J* = 7.4, 1.0 Hz, 11-C*H*), 7.83 (2H, dd, *J* = 8.7, 5.8 Hz, 17-C*H*), 7.10 (2H, s, 4-C*H*), 4.65 (6H, s, 19-CH_3), 2.88 (4H, t, *J* = 7.4 Hz, 7-CH_2), 2.28 (6H, s, 9-CH_3), 2.12 (2H, p, *J* = 7.4 Hz, 8-CH_2).

^{13}C NMR (126 MHz, D_2O) δ 149.0 (18-*C*H), 145.3 (16-*C*H), 139.5 (14-*C*), 138.5 (3-*C*H or 2-*C*H), 137.1 (3-*C*H or 2-*C*H), 135.6 (6-*C*), 135.2 (10-*C*), 135.0 (12-*C*H), 133.4 (5-*C*), 131.5 (4-*C*H), 130.8 (11-C*H*), 128.1 (15-*C*), 121.2 (17-*C*H), 117.5 (13-*C*H), 45.8 (19-*C*H_3), 37.4 (7-*C*H_2), 22.9 (8-*C*H_2), 13.4 (9-*C*H_3).

ν_{max} / cm^{-1} (compressed solid) 3042 (w), 3000 (w), 2942 (w), 2916 (w), 2840 (w), 1614, m), 1586 (s), 1531 (s), 1477 (w), 1463 (w), 1355 (m), 1330 (m), 1229 (m), 1156 (m), 1133 (w), 1117 (w), 1085 (w), 1070 (w), 993 (w), 978 (w), 855 (w), 795 (s), 762 (m), 721 (w), 693 (w), 657 (w), 574 (m).

ESI-HRMS for $C_{35}H_{32}N_2S_2^{+}$ $[M]^{2+}$ calcd: 272.1003, found: 272.1001.

References

1 M. P. O'Hagan, J. C. Morales and M. C. Galan, *European J. Org. Chem.*, 2019, 4995–5017.

2 A. Rich, *Gene*, 1993, **135**, 99–109.

3 B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell*, Garland Science, New York, 6th edn., 2015, p. 173.

4 B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell*, Garland Science, New York, 6th edn., 2015, p. 340.

5 S. T. G. Street, PhD Thesis, University of Bristol, 2017.

6 H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci.*, 1981, **78**, 2179–2183.

7 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.

8 A. Rich, *Nat. Struct. Mol. Biol.*, 2003, **10**, 247–249.

9 X.-J. Lu, Z. Shakked and W. K. Olson, *J. Mol. Biol.*, 2000, **300**, 819–840.

10 M. McCall, T. Brown and O. Kennard, *J. Mol. Biol.*, 1985, **183**, 385–396.

11 A. Rich and S. Zhang, *Nat. Rev. Genet.*, 2003, **4**, 566–572.

12 A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680–686.

13 J. Donohue, *Proc. Natl. Acad. Sci.*, 1953, **39**, 470–478.

14 K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 907–916.

15 E. N. Nikolova, H. Zhou, F. L. Gottardo, H. S. Alvey, I. J. Kimsey and H. M. Al-Hashimi, *Biopolymers*, 2013, **99**, 955–968.

16 J. L. Asensio, T. Brown and A. N. Lane, *Structure*, 1999, **7**, 1–11.

17 I. Bang, *Biochem. Z.*, 1910, **26**, 293–311.

18 M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013–2018.

19 C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps and A. Pluckthun, *Proc. Natl. Acad. Sci.*, 2001, **98**, 8572–8577.

20 S. Arnott, R. Chandrasekaran and C. M. Marttila, *Biochem. J.*, 1974, **141**, 537–43.

21 S. B. Zimmerman, G. H. Cohen and D. R. Davies, *J. Mol. Biol.*, 1975, **92**, 181–192.

22 E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco and E. H. Blackburn, *Cell*, 1987, **51**, 899–908.

23 D. Sen and W. Gilbert, *Nature*, 1988, **334**, 364–366.

24 G. Laughlan, A. Murchie, D. Norman, M. Moore, P. Moody, D. Lilley and B. Luisi,

Science (80-.)., 1994, **265**, 520–524.

25 F. Aboul-Ela, A. I. H. Murchie and D. M. J. Lilley, *Nature*, 1992, **360**, 280–282.

26 F. W. Smith and J. Feigon, *Nature*, 1992, **356**, 164–168.

27 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.

28 Y. Wang and D. J. Patel, *Biochemistry*, 1992, **31**, 8112–8119.

29 Y. Wang and D. J. Patel, *J. Mol. Biol.*, 1993, **234**, 1171–1183.

30 C. Kang, X. Zhang, R. Ratliff, R. Moyzis and A. Rich, *Nature*, 1992, **356**, 126–131.

31 R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sci.*, 1993, **90**, 3745–3749.

32 J. Dai, M. Carver and D. Yang, *Biochimie*, 2008, **90**, 1172–1183.

33 S. Neidle, *Therapeutic Applications of Quadruplex Nucleic Acids*, Elsevier, London, 2012.

34 H. Yaku, T. Fujimoto, T. Murashima, D. Miyoshi and N. Sugimoto, *Chem. Commun.*, 2012, **48**, 6203.

35 R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res.*, 2011, **39**, 5768–5775.

36 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.

37 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963–9970.

38 N. Zhang, A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2005, **127**, 17277–17285.

39 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, 5, 182–186.

40 G. F. Salgado, C. Cazenave, A. Kerkour and J. L. Mergny, *Chem. Sci.*, 2015, **6**, 3314–3320.

41 N. Maizels and L. T. Gray, *PLoS Genet.*, 2013, **9**, e1003468.

42 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.

43 R. Hänsel-Hertsch, M. Di Antonio and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2017, 18, 279–284.

44 S. Neidle, *J. Med. Chem.*, 2016, **59**, 5987–6011.

45 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261–275.

46 J.-L. Mergny and D. Sen, *Chem. Rev.*, 2019, **119**, 6290–6325.

47 F. Wang, X. Liu and I. Willner, *Angew. Chemie - Int. Ed.*, 2015, **54**, 1098–1129.

48 V. Yadav, Hemansi, N. Kim, N. Tuteja and P. Yadav, *Front. Plant Sci.*, 2017, **8**, 1–9.

49 S. K. Mishra, N. Jain, U. Shankar, A. Tawani, T. K. Sharma and A. Kumar, *Sci. Rep.*, 2019, **9**, 1791.

50 M. Metifiot, S. Amrane, S. Litvak and M. L. Andreola, *Nucleic Acids Res.*, 2014, **42**, 12352–12366.

51 S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M.-L. Andreola and J.-L. Mergny, *J. Am. Chem. Soc.*, 2014, **136**, 5249–5252.

52 E. Belmonte-Reche, M. Martínez-García, A. Guédin, M. Zuffo, M. Arévalo-Ruiz, F. Doria, J. Campos-Salinas, M. Maynadier, J. J. López-Rubio, M. Freccero, J.-L. Mergny, J. M. Pérez-Victoria and J. C. Morales, *J. Med. Chem.*, 2018, **61**, 1231–1240.

53 L. M. Harris and C. J. Merrick, *PLOS Pathog.*, 2015, **11**, e1004562.

54 E. Bottius, N. Bakhsis and A. Scherf, *Mol. Cell. Biol.*, 2015, **18**, 919–925.

55 E. Puig Lombardi and A. Londoño-Vallejo, *Nucleic Acids Res.*, 2020, **48**, 1–15.

56 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908–2916.

57 A. Bedrat, L. Lacroix and J.-L. L. Mergny, *Nucleic Acids Res.*, 2016, **44**, 1746–1759.

58 T. Ou, Y. Lu, J. Tan, Z. Huang, K.-Y. Wong and L. Gu, *ChemMedChem*, 2008, **3**, 690–713.

59 O. Tacar, P. Sriamornsak and C. R. Dass, *J. Pharm. Pharmacol.*, 2013, **65**, 157–170.

60 D. Hanahan and R. A. Weinberg, *Cell*, 2011, **144**, 646–674.

61 J. A. Hackett, D. M. Feldser and C. W. Greider, *Cell*, 2001, **106**, 275–286.

62 J. Meyne, R. L. Ratliff and R. K. Moyzis, *Proc. Natl. Acad. Sci.*, 1989, 86, 7049–7053.

63 R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff and J. R. Wu, *Proc. Natl. Acad. Sci.*, 1988, **85**, 6622–6626.

64 T. De Lange, *Genes Dev.*, 2005, **19**, 2100–2110.

65 W. Palm and T. de Lange, *Annu. Rev. Genet.*, 2008, **42**, 301–334.

66 C. B. Harley, A. B. Futcher and C. W. Greider, *Nature*, 1990, 345, 458–460.

67 R. Ohki, T. Tsurimoto and F. Ishikawa, *Mol. Cell. Biol.*, 2001, **21**, 5753–5766.

68 X. Zhang, V. Mar, W. Zhou, L. Harrington and M. O. Robinson, *Genes Dev.*, 1999, **13**, 2388–2399.

69 A. G. Bodnar, *Science (80-.).*, 1998, **279**, 349–352.

70 C. W. Greider and E. H. Blackburn, *Cell*, 1985, **43**, 405–413.

71 C. M. Counter, A. A. Avilion, C. E. LeFeuvre, N. G. Stewart, C. W. Greider, C. B. Harley and S. Bacchetti, *EMBO J.*, 1992, **11**, 1921–1929.

72 J. W. Shay and W. E. Wright, *Semin. Cancer Biol.*, 2011, **21**, 349–353.

73 J.-L. Mergny, *Nucleic Acids Res.*, 2002, **30**, 839–865.

74 D. E. Gomez, A. M. Tejera and O. A. Olivero, *Biochem. Biophys. Res. Commun.*, 1998, **246**, 107–110.

75 E. Pascolo, C. Wenz, J. Lingner, N. Hauel, H. Priepke, I. Kauffmann, P. Garin-Chesa, W. J. Rettig, K. Damm and A. Schnapp, *J. Biol. Chem.*, 2002, **277**, 15566–15572.

76 A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott, *Nature*, 1991, **350**, 718–720.

77 D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle and L. H. Hurley, *J. Med. Chem.*, 1997, **40**, 2113–2116.

78 S. Neidle, *FEBS J.*, 2010, **277**, 1118–1125.

79 N. S. Ilyinsky, A. M. Varizhuk, A. D. Beniaminov, M. A. Puzanov, A. K. Shchyolkina and D. N. Kaluzhny, *Mol. Biol.*, 2014, **48**, 778–794.

80 J. T. Davis, *Angew. Chemie - Int. Ed.*, 2004, **43**, 668–698.

81 S. A. Ohnmacht and S. Neidle, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2602–2612.

82 L. Parrotta, F. Ortuso, F. Moraca, R. Rocca, G. Costa, S. Alcaro and A. Artese, *Expert Opin. Drug Discov.*, 2014, **9**, 1167–1187.

83 A. R. Duarte, E. Cadoni, A. S. Ressurreição, R. Moreira and A. Paulo, *ChemMedChem*, 2018, **13**, 869–893.

84 S. Asamitsu, S. Obata, Z. Yu, T. Bando and H. Sugiyama, *Molecules*, 2019, **24**, 429.

85 L. Sabater, M.-L. Nicolau-Travers, A. De Rache, E. Prado, J. Dejeu, O. Bombarde, J. Lacroix, P. Calsou, E. Defrancq, J.-L. Mergny, D. Gomez and G. Pratviel, *JBIC J. Biol. Inorg. Chem.*, 2015, **20**, 729–738.

86 D. Gomez, M.-F. O'Donohue, T. Wenner, C. Douarre, J. Macadré, P. Koebel, M.-J. Giraud-Panis, H. Kaplan, A. Kolkes, K. Shin-ya and J.-F. Riou, *Cancer Res.*, 2006, **66**, 6908–6912.

87 F. d'Adda di Fagagna, *Nat. Rev. Cancer*, 2008, **8**, 512–522.

88 F. Dietlein, L. Thelen and H. C. Reinhardt, *Trends Genet.*, 2014, **30**, 326–339.

89 J.-B. Vannier, V. Pavicic-Kaltenbrunner, M. I. R. Petalcorin, H. Ding and S. J. Boulton, *Cell*, 2012, **149**, 795–806.

90 V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol.*, 2015, **33**, 877–881.

91 R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill and S. Balasubramanian, *Nat. Genet.*, 2016, **48**, 1267–1272.

92 A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci.*, 2002, **99**, 11593–11598.

93 P. Agrawal, C. Lin, R. I. Mathad, M. Carver and D. Yang, *J. Am. Chem. Soc.*, 2014, **136**, 1750–1753.

94 A. Membrino, S. Cogoi, E. B. Pedersen and L. E. Xodo, *PLoS One*, 2011, **6**, e24421.

95 S. Cogoi, M. Paramasivam, B. Spolaore and L. E. Xodo, *Nucleic Acids Res.*, 2008, **36**, 3765–3780.

96 S. T. D. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 13399–13409.

97 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2007, **35**, 406–413.

98 M. C. Field, D. Horn, A. H. Fairlamb, M. A. J. Ferguson, D. W. Gray, K. D. Read, M. De Rycker, L. S. Torrie, P. G. Wyatt, S. Wyllie and I. H. Gilbert, *Nat. Rev. Microbiol.*, 2017, **15**, 217–231.

99 J. R. Franco, G. Cecchi, G. Priotto, M. Paone, A. Diarra, L. Grout, R. C. Mattioli and D. Argaw, *PLoS Negl. Trop. Dis.*, 2017, **11**, 1–26.

100 M. P. Barrett, R. J. S. Burchmore, A. Stich, J. O. Lazzari, A. C. Frasch, J. J. Cazzulo and S. Krishna, *Lancet*, 2003, **362**, 1469–1480.

101 A. Luscher, H. de Koning and P. Maser, *Curr. Pharm. Des.*, 2007, **13**, 555–567.

102 N. Baker, H. P. de Koning, P. Mäser and D. Horn, *Trends Parasitol.*, 2013, **29**, 110–118.

103 M. Zuffo, A. Stucchi, J. Campos-Salinas, M. Cabello-Donayre, M. Martínez-García, E. Belmonte-Reche, J. M. Pérez-Victoria, J. L. Mergny, M. Freccero, J. C. Morales and F. Doria, *Eur. J. Med. Chem.*, 2019, **163**, 54–66.

104 D. Monchaud and M.-P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627–636.

105 E. Ruggiero and S. N. Richter, *Nucleic Acids Res.*, 2018, **46**, 3270–3283.

106 J. Dash, P. S. Shirude, S. T. D. Hsu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15950–15956.

107 M. Bončina, Č. Podlipnik, I. Piantanida, J. Eilmes, M. P. Teulade-Fichou, G. Vesnaver and J. Lah, *Nucleic Acids Res.*, 2015, **43**, 10376–10386.

108 A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, P. Alberti, M. P. Teulade-Fichou, L. Lacroix and J. L. Mergny, *Methods*, 2007, **42**, 183–195.

109 R. I. Mathad and D. Yang, in *Methods in molecular biology (Clifton, N.J.)*, 2011, pp. 77–96.

110 N. H. Campbell, G. N. Parkinson, A. P. Reszka and S. Neidle, *J. Am. Chem. Soc.*, 2008, **130**, 6722–6724.

111 R. del Villar-Guerra, J. O. Trent and J. B. Chaires, *Angew. Chemie - Int. Ed.*, 2018, **57**, 7171–7175.

112 H. Xin, *Methods Mol. Biol.*, 2011, **735**, 107–111.

113 P. J. Perry, S. M. Gowan, A. P. Reszka, P. Polucci, T. C. Jenkins, L. R. Kelland and S. Neidle, *J. Med. Chem.*, 1998, **41**, 3253–3260.

114 P. J. Perry, A. P. Reszka, A. A. Wood, M. A. Read, S. M. Gowan, H. S. Dosanjh, J. O. Trent, T. C. Jenkins, L. R. Kelland and S. Neidle, *J. Med. Chem.*, 1998, **41**, 4873–4884.

115 P. J. Perry, M. A. Read, R. T. Davies, S. M. Gowan, A. P. Reszka, A. A. Wood, L. R.

Kelland and S. Neidle, *J. Med. Chem.*, 1999, **42**, 2679–2684.

116 R. J. Harrison, S. M. Gowan, L. R. Kelland and S. Neidle, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2463–2468.

117 M. J. B. Moore, C. M. Schultes, J. Cuesta, F. Cuenca, M. Gunaratnam, F. A. Tanious, W. D. Wilson and S. Neidle, *J. Med. Chem.*, 2006, **49**, 582–599.

118 A. M. Burger, F. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double and S. Neidle, *Cancer Res.*, 2005, **65**, 1489–1496.

119 O. Y. Fedoroff, M. Salazar, H. Han, V. V. Chemeris, S. M. Kerwin and L. H. Hurley, *Biochemistry*, 1998, **37**, 12367–12374.

120 J. T. Kern and S. M. Kerwin, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 3395–3398.

121 F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.

122 M. Micco, G. W. Collie, A. G. Dale, S. A. Ohnmacht, I. Pazitna, M. Gunaratnam, A. P. Reszka and S. Neidle, *J. Med. Chem.*, 2013, **56**, 2959–2974.

123 D. M. Răsădean, B. Sheng, J. Dash and G. D. Pantoş, *Chem. Eur. J.*, 2017, **23**, 8491–8499.

124 S. T. G. Street, D. N. Chin, G. J. Hollingworth, M. Berry, J. C. Morales and M. C. Galan, *Chem. Eur. J.*, 2017, **23**, 6953–6958.

125 M. Arévalo-Ruiz, F. Doria, E. Belmonte-Reche, A. De Rache, J. Campos-Salinas, R. Lucas, E. Falomir, M. Carda, J. M. Pérez-Victoria, J. L. Mergny, M. Freccero and J. C. Morales, *Chem. Eur. J.*, 2017, **23**, 2157–2164.

126 M. Zuffo, F. Doria, S. Botti, G. Bergamaschi and M. Freccero, *Biochim. Biophys. Acta - Gen. Subj.*, 2017, **1861**, 1303–1311.

127 G. W. Collie, R. Promontorio, S. M. Hampel, M. Micco, S. Neidle and G. N. Parkinson, *J. Am. Chem. Soc.*, 2012, **134**, 2723–2731.

128 S. A. Ohnmacht, C. Marchetti, M. Gunaratnam, R. J. Besser, S. M. Haider, G. Di Vita, H. L. Lowe, M. Mellinas-Gomez, S. Diocou, M. Robson, J. Šponer, B. Islam, R. Barbara Pedley, J. A. Hartley and S. Neidle, *Sci. Rep.*, 2015, **5**, 11385.

129 J. Ren and J. B. Chaires, *Biochemistry*, 1999, **38**, 16067–16075.

130 A. De Cian, E. DeLemos, J. L. Mergny, M. P. Teulade-Fichou and D. Monchaud, *J. Am. Chem. Soc.*, 2007, **129**, 1856–1857.

131 R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J. F. Riou and S. Balasubramanian, *J. Am. Chem. Soc.*, 2008, **130**, 15758–15759.

132 K. Shin-ya, K. Wierzba, K. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa and H. Seto, *J. Am. Chem. Soc.*, 2001, **123**, 1262–1263.

133 W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa and A. T. Phan, *J. Am. Chem. Soc.*, 2013, **135**, 13495–13501.

134 A. Nakajima, T. Tauchi, G. Sashida, M. Sumi, K. Abe, K. Yamamoto, J. H. Ohyashiki and K. Ohyashiki, *Leukemia*, 2003, **17**, 560–567.

135 T. Doi, M. Yoshida, K. Shin-Ya and T. Takahashi, *Org. Lett.*, 2006, **8**, 4165–4167.

136 S. N. Georgiades, N. H. Abd Karim, K. Suntharalingam and R. Vilar, *Angew. Chemie Int. Ed.*, 2010, **49**, 4020–4034.

137 S. Bandeira, J. Gonzalez-Garcia, E. Pensa, T. Albrecht and R. Vilar, *Angew. Chemie Int. Ed.*, 2018, **57**, 310–313.

138 P. Alberti, A. Bourdoncle, B. Saccà, L. Lacroix and J. L. Mergny, *Org. Biomol. Chem.*, 2006, **4**, 3383–3391.

139 W. U. Dittmer, A. Reuter and F. C. Simmel, *Angew. Chemie Int. Ed.*, 2004, **43**, 3550–3553.

140 T. Li, S. Dong and E. Wang, *J. Am. Chem. Soc.*, 2010, **132**, 13156–13157.

141 P. Alberti and J.-L. Mergny, *Proc. Natl. Acad. Sci.*, 2003, **100**, 1569–1573.

142 J. S. Kahn, A. Trifonov, A. Cecconello, W. Guo, C. Fan and I. Willner, *Nano Lett.*, 2015, **15**, 7773–7778.

143 X. Wang, J. Huang, Y. Zhou, S. Yan, X. Weng, X. Wu, M. Deng and X. Zhou, *Angew. Chemie - Int. Ed.*, 2010, **49**, 5305–5309.

144 R. Rodriguez, G. D. Pantoş, D. P. N. Gonçalves, J. K. M. Sanders and S. Balasubramanian, *Angew. Chemie - Int. Ed.*, 2007, **46**, 5405–5407.

145 T. Tian, Y. Song, L. Wei, J. Wang, B. Fu, Z. He, X. R. Yang, F. Wu, G. Xu, S. M. Liu, C. Li, S. Wang and X. Zhou, *Nucleic Acids Res.*, 2017, **45**, 2283–2293.

146 S. A. Kovalenko, A. L. Dobryakov, I. Ioffe and N. P. Ernsting, *Chem. Phys. Lett.*, 2010, **493**, 255–258.

147 G. Mayer and A. Hechel, *Angew. Chemie - Int. Ed.*, 2006, **45**, 4900–4921.

148 W. A. Velema, W. Szymanski and B. L. Feringa, *J. Am. Chem. Soc.*, 2014, **136**, 2178–2191.

149 A. S. Lubbe, W. Szymanski and B. L. Feringa, *Chem. Soc. Rev.*, 2017, **46**, 1052–1079.

150 A. Gautier, C. Gauron, M. Volovitch, D. Bensimon, L. Jullien and S. Vriz, *Nat. Chem. Biol.*, 2014, **10**, 533–541.

151 W. Szymański, J. M. Beierle, H. A. V. Kistemaker, W. A. Velema and B. L. Feringa, *Chem. Rev.*, 2013, **113**, 6114–6178.

152 L. Wu, K. Koumoto and N. Sugimoto, *Chem. Commun.*, 2009, 1915–1917.

153 S. Ogasawara and M. Maeda, *Angew. Chemie - Int. Ed.*, 2009, **48**, 6671–6674.

154 J. Thevarpadam, I. Bessi, O. Binas, D. P. N. Gonçalves, C. Slavov, H. R. A. Jonker, C. Richter, J. Wachtveitl, H. Schwalbe and A. Heckel, *Angew. Chemie - Int. Ed.*, 2016, **55**, 2738–2742.

155 J. H. R. Tucker, A. Ali, G. Bullen, B. Cross, T. Dafforn, H. A. Little, J. Manchester and A. Peacock, *Chem. Commun.*, 2019, 2–5.

156 M. M. Lerch, M. J. Hansen, G. M. van Dam, W. Szymanski and B. L. Feringa, *Angew. Chemie - Int. Ed.*, 2016, **55**, 10978–10999.

157 J. Broichhagen, J. A. Frank and D. Trauner, *Acc. Chem. Res.*, 2015, **48**, 1947–1960.

158 K. Hüll, J. Morstein and D. Trauner, *Chem. Rev.*, 2018, **118**, 10710–10747.

159 W. Szymanski, M. E. Ourailidou, W. A. Velema, F. J. Dekker and B. L. Feringa, *Chem. - A Eur. J.*, 2015, **21**, 16517–16524.

160 J. Andersson, S. Li, P. Lincoln and J. Andréasson, *J. Am. Chem. Soc.*, 2008, **130**, 11836–11837.

161 J. R. Nilsson, S. Li, B. Önfelt and J. Andréasson, *Chem. Commun.*, 2011, **47**, 11020.

162 P. Murat, M. V. Gormally, D. Sanders, M. Di Antonio and S. Balasubramanian, *Chem. Commun.*, 2013, **49**, 8453–8455.

163 T. Nakamura, K. Iida, M. Tera, K. Shin-ya, H. Seimiya and K. Nagasawa, *ChemBioChem*, 2012, **13**, 774–777.

164 T. Tian, Y. Song, J. Wang, B. Fu, Z. He, X. Xu, A. Li, X. Zhou, S. Wang and X. Zhou, *J. Am. Chem. Soc.*, 2016, **138**, 955–961.

165 X. Xing, X. Wang, L. Xu, Y. Tai, L. Dai, X. Zheng, W. Mao, X. Xu and X. Zhou, *Org. Biomol. Chem.*, 2011, **9**, 6639–6645.

166 M. Dudek, M. Deiana, Z. Pokladek, P. Mlynarz, M. Samoc and K. Matczyszyn, *Nanoscale*, 2018, **10**, 11302–11306.

167 J. E. B. McCallum, C. W. Coyle, R. R. Elson and B. A. Titterington, *Nucleosides, Nucleotides and Nucleic Acids*, 2018, **37**, 169–178.

168 A. Bergen, S. Rudiuk, M. Morel, T. Le Saux, H. Ihmels and D. Baigl, *Nano Lett.*, 2016, **16**, 773–780.

169 M. P. O'Hagan, P. Peñalver, R. S. L. Gibson, J. C. Morales and M. C. Galan, *Chem. – A Eur. J.*, 2020, chem.201905753.

170 E. M. M. Tan, S. Amirjalayer, S. Smolarek, A. Vdovin, F. Zerbetto and W. J. Buma, *Nat. Commun.*, 2015, **6**, 5860.

171 I. V. Komarov, S. Afonin, O. Babii, T. Schober and A. S. Ulrich, *Chem. - A Eur. J.*, 2018, **24**, 11245–11254.

172 D. Villarón and S. Wezenberg, *Angew. Chemie Int. Ed.*, , DOI:10.1002/anie.202001031.

173 J. E. McMurry, M. P. Fleming, K. L. Kees and L. R. Krepski, *J. Org. Chem.*, 1978, **43**, 3255–3266.

174 T. Mukaiyama, T. Sato and J. Hanna, *Chem. Lett.*, 1973, **2**, 1041–1044.

175 K. Ogawa, J. Harada and S. Tomoda, *Acta Crystallogr. Sect. B Struct. Sci.*, 1995, **51**,

240–248.

176 M. Oelgemöller, B. Brem, R. Frank, S. Schneider, D. Lenoir, N. Hertkorn, Y. Origane, P. Lemmen, J. Lex and Y. Inoue, *J. Chem. Soc., Perkin Trans. 2*, 2002, 1760–1771.

177 S. J. Wezenberg, M. Vlatković, J. C. M. Kistemaker and B. L. Feringa, *J. Am. Chem. Soc.*, 2014, **136**, 16784–16787.

178 S. J. Wezenberg and B. L. Feringa, *Org. Lett.*, 2017, **19**, 324–327.

179 J. Jong, B. L. Feringa and S. J. Wezenberg, *ChemPhysChem*, 2019, **20**, 3306–3310.

180 T. Shimasaki, S. Kato, K. Ideta, K. Goto and T. Shinmyozu, *J. Org. Chem.*, 2007, **72**, 1073–1087.

181 S. J. Wezenberg and B. L. Feringa, *Nat. Commun.*, 2018, **18**, 1984.

182 A. S. Lubbe, Q. Liu, S. J. Smith, J. W. De Vries, J. C. M. Kistemaker, A. H. De Vries, I. Faustino, Z. Meng, W. Szymanski, A. Herrmann and B. L. Feringa, *J. Am. Chem. Soc.*, 2018, **140**, 5069–5076.

183 F. Khalili, A. Henni and A. L. L. East, *J. Chem. Eng. Data*, 2009, **54**, 2914–2917.

184 B. P. Fors, D. A. Watson, M. R. Biscoe and S. L. Buchwald, *J. Am. Chem. Soc.*, 2008, **130**, 13552–13554.

185 M. M. Heravi, Z. Kheilkordi, V. Zadsirjan, M. Heydari and M. Malmir, *J. Organomet. Chem.*, 2018, **861**, 17–104.

186 P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 1305–1323.

187 A. De Rache and J. L. Mergny, *Biochimie*, 2015, **115**, 194–202.

188 H. Mach, D. B. Volkin, C. J. Burke and C. R. Middaugh, in *Protein Stability and Folding*, Humana Press, New Jersey, 1995, pp. 91–114.

189 A. Beer, *Ann. der Phys. und Chemie*, 1852, **162**, 78–88.

190 C. I. V. Ramos, S. P. Almeida, L. M. O. Lourenço, P. M. R. Pereira, R. Fernandes, M. F. Amparo Faustino, J. P. C. Tomé, J. Carvalho, C. Cruz and M. P. M. S. Graça Neves, *Molecules*, 2019, **24**, 1–23.

191 N. J. Greenfield, *Nat. Protoc.*, 2006, **1**, 2876–2890.

192 B. A. Wallace and R. W. Janes, *Modern techniques for circular dichroism and synchrotron radiation circular dichroism spectroscopy*, IOS Press, 2009.

193 S. Paramasivan, I. Rujan and P. H. Bolton, *Methods*, 2007, **43**, 324–331.

194 N. C. Garbett, P. A. Ragazzon and J. O. B. Chaires, *Nat. Protoc.*, 2007, **2**, 3166–3172.

195 A. Randazzo, G. P. Spada and M. W. da Silva, in *Topics in Current Chemistry*, Springer, Heidelberg, 2012, vol. 330, pp. 67–86.

196 Z. A. E. Waller, S. A. Sewitz, S. T. D. Hsu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 12628–12633.

197 T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma,

K.-Y. Wong, J. C.-O. Tang, A. S.-C. Chan and L.-Q. Gu, *J. Med. Chem.*, 2007, **50**, 1465–1474.

198 M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1–16.

199 I. R. Kleckner and M. P. Foster, *Biochim. Biophys. Acta - Proteins Proteomics*, 2011, **1814**, 942–968.

200 K. Schmidt-Rohr, *ACS Omega*, 2020, **5**, 2221–2233.

201 E. M. Larson, D. J. Doughman, D. S. Gregerson and W. F. Obritsch, *Investig. Ophthalmol. Vis. Sci.*, 1997, **38**, 1929–1933.

202 J. M. Pérez-Victoria, B. I. Bavchvarov, I. R. Torrecillas, M. Martínez-García, C. López-Martín, M. Campillo, S. Castanys and F. Gamarro, *Antimicrob. Agents Chemother.*, 2011, **55**, 3838–3844.

203 M. P. O'Hagan, S. Haldar, M. Duchi, T. A. A. Oliver, A. J. Mulholland, J. C. Morales and M. C. Galan, *Angew. Chemie Int. Ed.*, 2019, **58**, 4334–4338.

204 A. J. J. Lennox and G. C. Lloyd-Jones, *Chem. Soc. Rev.*, 2014, **43**, 412–443.

205 L. Hahn, N. J. Buurma and L. H. Gade, *Chem. - A Eur. J.*, 2016, **22**, 6314–6322.

206 F. Moraca, J. Amato, F. Ortuso, A. Artese, B. Pagano, E. Novellino, S. Alcaro, M. Parrinello and V. Limongelli, *Proc. Natl. Acad. Sci.*, 2017, **114**, E2136–E2145.

207 Z. F. Wang, M. H. Li, S. T. D. Hsu and T. C. Chang, *Nucleic Acids Res.*, 2014, **42**, 4723–4733.

208 R. D. Gray, J. Li and J. B. Chaires, *J. Phys. Chem. B*, 2009, **113**, 2676–2683.

209 O. S. Andersen, in *Encyclopedia of Metalloproteins*, eds. R. H. Kretsinger, V. N. Uversky and E. A. Permyakov, Springer, New York, 2013.

210 J. F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Megnin-Chanet, C. Helene and J. L. Mergny, *Proc. Natl. Acad. Sci.*, 2002, **99**, 2672–2677.

211 S. Zhang, H. Sun, H. Chen, Q. Li, A. Guan, L. Wang, Y. Shi, S. Xu, M. Liu and Y. Tang, *Biochim. Biophys. Acta - Gen. Subj.*, 2018, **1862**, 1101–1106.

212 C. S. Foote, *Science (80-.).*, 1968, **162**, 963–970.

213 C. D. Clark, J. D. Debad, E. H. Yonemoto, T. E. Mallouk and A. J. Bard, *J. Am. Chem. Soc.*, 1997, **119**, 10525–10531.

214 J.-X. Wang, L.-Y. Niu, P.-Z. Chen, Y.-Z. Chen, Q.-Z. Yang and R. Boulatov, *Chem. Commun.*, 2019, **55**, 7017–7020.

215 M. P. O'Hagan, J. Ramos-Soriano, S. Haldar, S. Sheikh, J. C. Morales, A. J. Mulholland and M. C. Galan, *Chem. Commun.*, 2020, **56**, 5186–5189.

216 M. Irie and M. Mohri, *J. Org. Chem.*, 1988, **53**, 803–808.

217 M. Irie, *Chem. Rev.*, 2000, **100**, 1685–1716.

218 B. Roubinet, M. L. Bossi, P. Alt, M. Leutenegger, H. Shojaei, S. Schnorrenberg, S.

Nizamov, M. Irie, V. N. Belov and S. W. Hell, *Angew. Chemie - Int. Ed.*, 2016, **55**, 15429–15433.

219 T. Fukaminato, T. Hirose, T. Doi, M. Hazama, K. Matsuda and M. Irie, *J. Am. Chem. Soc.*, 2014, **136**, 17145–17154.

220 M. Herder, B. M. Schmidt, L. Grubert, M. Pätzel, J. Schwarz and S. Hecht, *J. Am. Chem. Soc.*, 2015, **137**, 2738–2747.

221 A. Mammana, G. T. Carroll, J. Areephong and B. L. Feringa, *J. Phys. Chem. B*, 2011, **115**, 11581–11587.

222 T. C. S. Pace, V. Müller, S. Li, P. Lincoln and J. Andréasson, *Angew. Chemie - Int. Ed.*, 2013, **52**, 4393–4396.

223 A. Presa, R. F. Brissos, A. B. Caballero, I. Borilovic, L. Korrodi-Gregório, R. Pérez-Tomás, O. Roubeau and P. Gamez, *Angew. Chemie - Int. Ed.*, 2015, **54**, 4561–4565.

224 F. Hu, L. Jiang, M. Cao, Z. Xu, J. Huang, D. Wu, W. Yang, S. H. Liu and J. Yin, *RSC Adv.*, 2015, **5**, 5982–5987.

225 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, **31**, NA-NA.

226 B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

227 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

228 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.

229 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

230 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, L. J. William, C. Jayaraman, D. M. Jeffry, W. I. Roger, L. K. Michael, W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

231 J. Åqvist, *J. Phys. Chem.*, 1990, **94**, 8021–8024.

232 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

233 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 2012, **3684**, 926–935.

234 S. Nosé, *Mol. Phys.*, 2002, **100**, 191–198.

235 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

236 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

237 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys.*

Commun., 2014, **185**, 604–613.

238 A. B. Pangborn, M. A. Giardello, R. H. Grubbs, R. K. Rosen and F. J. Timmers, *Organometallics*, 1996, **15**, 1518–1520.

239 W. C. Still, M. Kahn and A. Mitra, *J. Org. Chem.*, 1978, **43**, 2923–2925.

240 J. Park, D. Feng, S. Yuan and H.-C. Zhou, *Angew. Chemie Int. Ed.*, 2015, **54**, 430–435.

241 P. Remón, D. González, S. Li, N. Basílio, J. Andréasson and U. Pischel, *Chem. Commun.*, 2019, **55**, 4335–4338.

Appendix

A1 Derivation of 1:1 binding model

Beginning with Eqn (2.5) introduced in Chapter 2:

$$A_{[D]} = \varepsilon_L[L] + \varepsilon_{D.L}[D.L] \tag{2.5}$$

Firstly, it is noted that the free ligand concentration is simply the total ligand concentration minus the concentration of bound ligand:

$$[L] = [L]_{tot} - [D.L] \tag{A1}$$

Eqn. (A1) is substituted into Eqn. (2.5) to give Eqn. (A2).

$$A_{[D]} = \varepsilon_L([L]_{tot} - [D.L]) + \varepsilon_{D.L}[D.L] \tag{A2}$$

At the start of the titration only ligand is present and therefore the concentration of complex ($[D.L]$) is zero. The absorbance, $A_{[D]=0}$, is simply:

$$A_{[D]=0} = \varepsilon_L[L]_{tot} \tag{A3}$$

Subtracting Eqn. (A3) from Eqn. (A2) gives:

$$A_{[D]} - A_{[D]=0} = \varepsilon_L([L]_{tot} - [D.L]) + \varepsilon_{D.L}[D.L] - \varepsilon_L[L]_{tot} \tag{A4}$$

Which is simplified to:

$$\Delta A = \varepsilon_{\Delta D.L} \cdot [D.L] \tag{A5}$$

Where $\varepsilon_{\Delta D.L}$ is the difference in extinction coefficient between the bound complex and free ligand and the observed absorbance change is related directly to the concentration of the bound ligand. Eqn. (A5) is next expressed in terms of the association constant and known concentrations of the total DNA and ligand through the following argument:
By noting again Eqn. (A1) and also that:

$$[D] = [D]_{tot} - [D.L] \tag{A6}$$

The association constant, Eqn (2.3), can be expressed in terms of the concentration of bound complex and the known total concentrations of DNA and ligand:

$$K_a = \frac{[D.L]}{([D]_{tot} - [D.L]) \cdot ([L]_{tot} - [D.L])} \tag{A7}$$

Eqn. (A7) is then trivially rearranged and factorised into standard quadratic form:

$$[D.L]^2 - \left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right)[D.L] + [D]_{tot}[L]_{tot} = 0 \tag{A8}$$

For which the solution for $[D.L]$ is given by the quadratic formula:

$$[D.L] = \frac{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right) - \sqrt{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right)^2 - 4 \cdot [D]_{tot}[L]_{tot}}}{2} \tag{A9}$$

Finally, substitution of Eqn. (A9) into Eqn. (A5) directly relates the observed absorbance change to the known quantities $[D]_{tot}$ and $[L]_{tot}$ and the unknown quantities $\varepsilon_{\Delta D.L}$ and the association constant K_a to give Eqn. (2.6) as presented in Chapter 2:

$$\Delta A = \varepsilon_{\Delta D.L} \cdot \frac{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right) - \sqrt{\left([D]_{tot} + [L]_{tot} + \frac{1}{K_a}\right)^2 - 4 \cdot [D]_{tot}[L]_{tot}}}{2} \tag{2.6}$$

A2 Supplementary data for Chapter 2

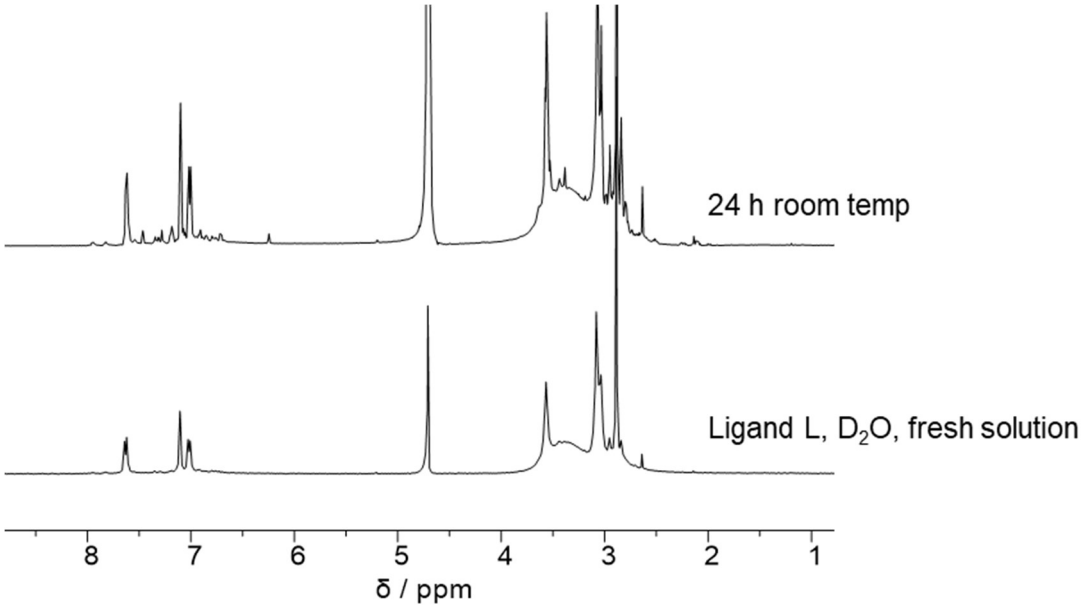


Figure A2.1: Decomposition of ligand **L** in D_2O solution. Similar results were obtained for ligands **M** and **N**.

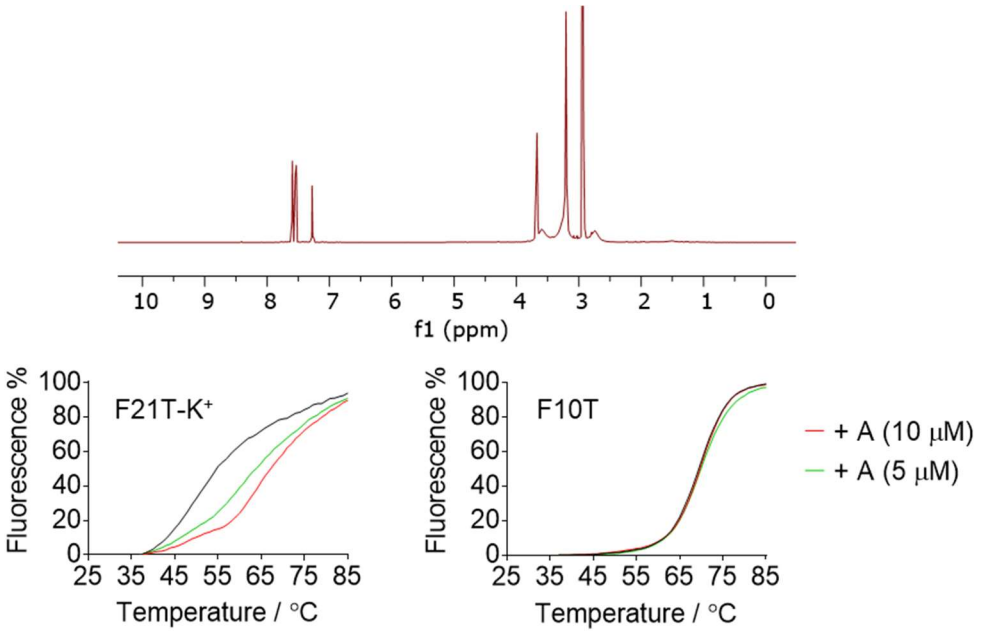


Figure A2.2: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **A**

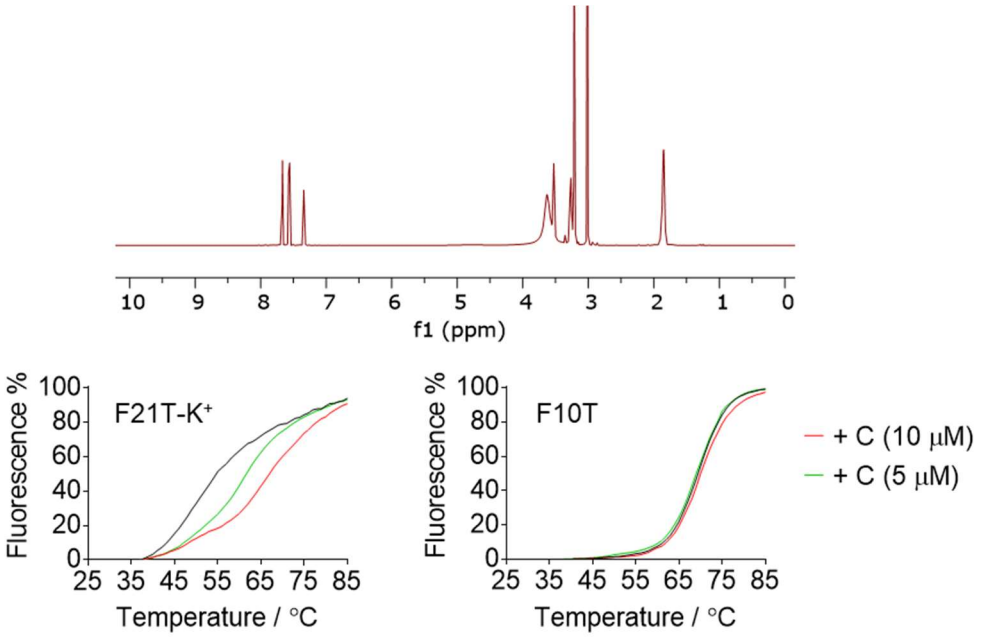


Figure A2.3: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **C**

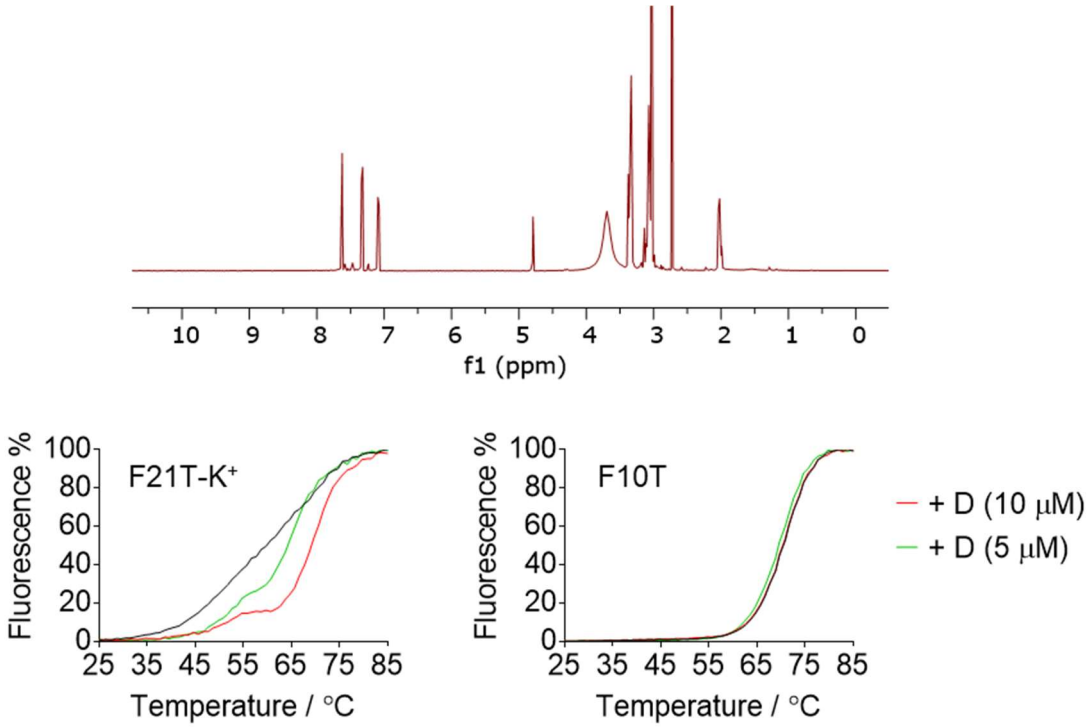


Figure A2.4: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **D**

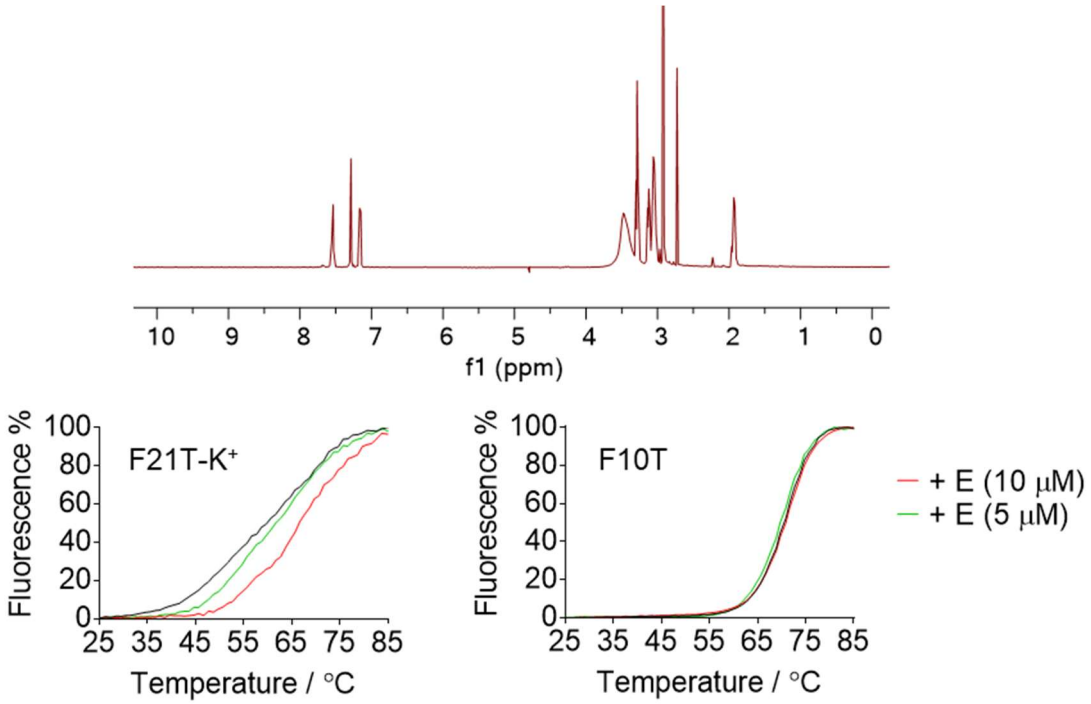


Figure A2.5: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **E**

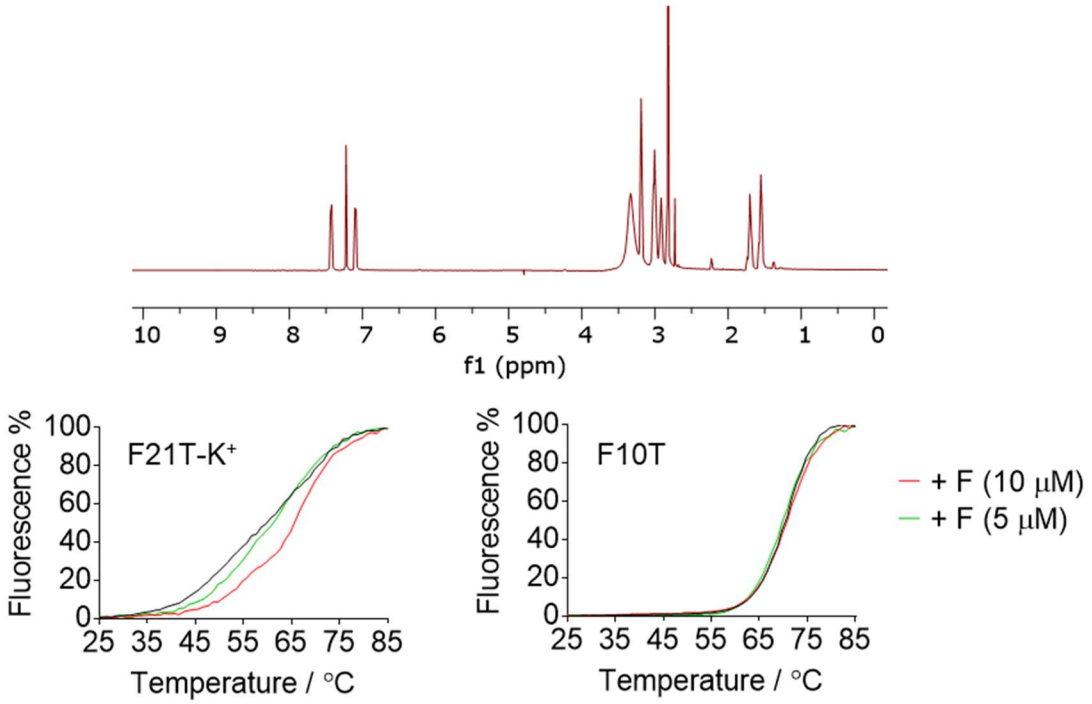


Figure A2.6: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **F**

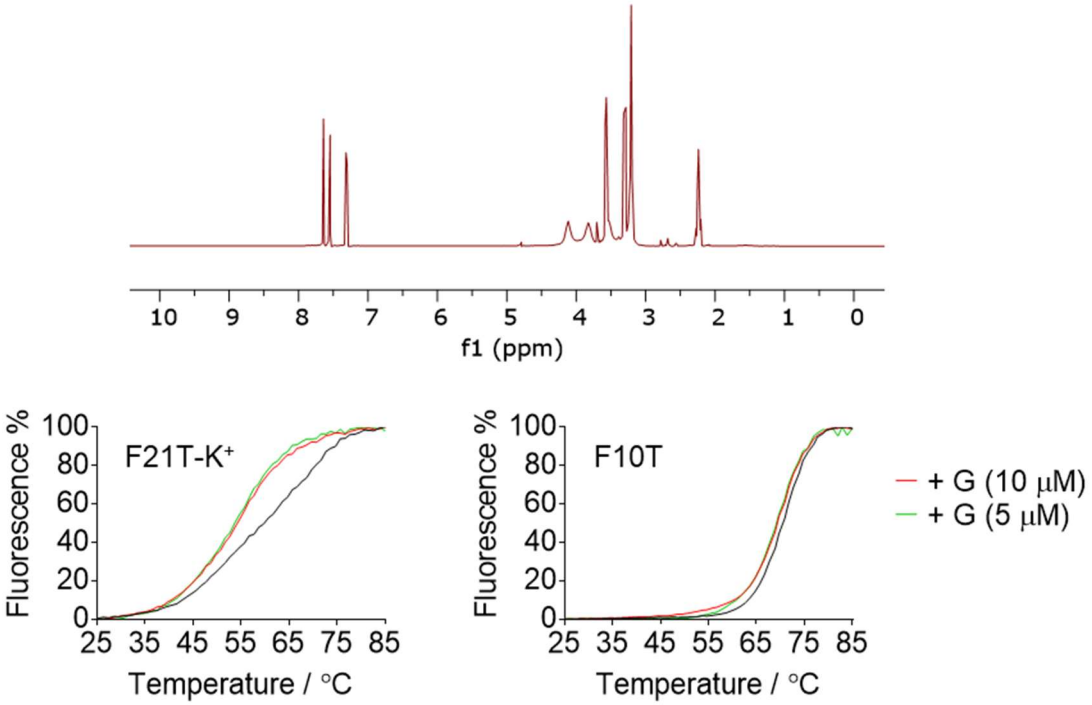


Figure A2.7: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **G**

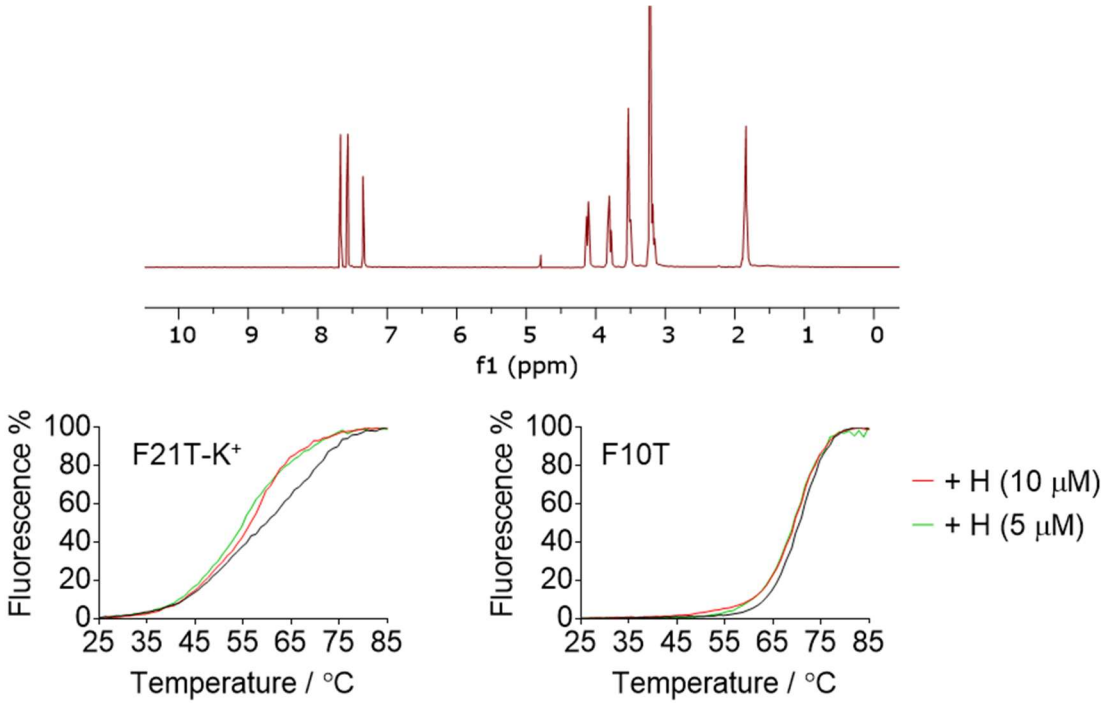


Figure A2.8: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **H**

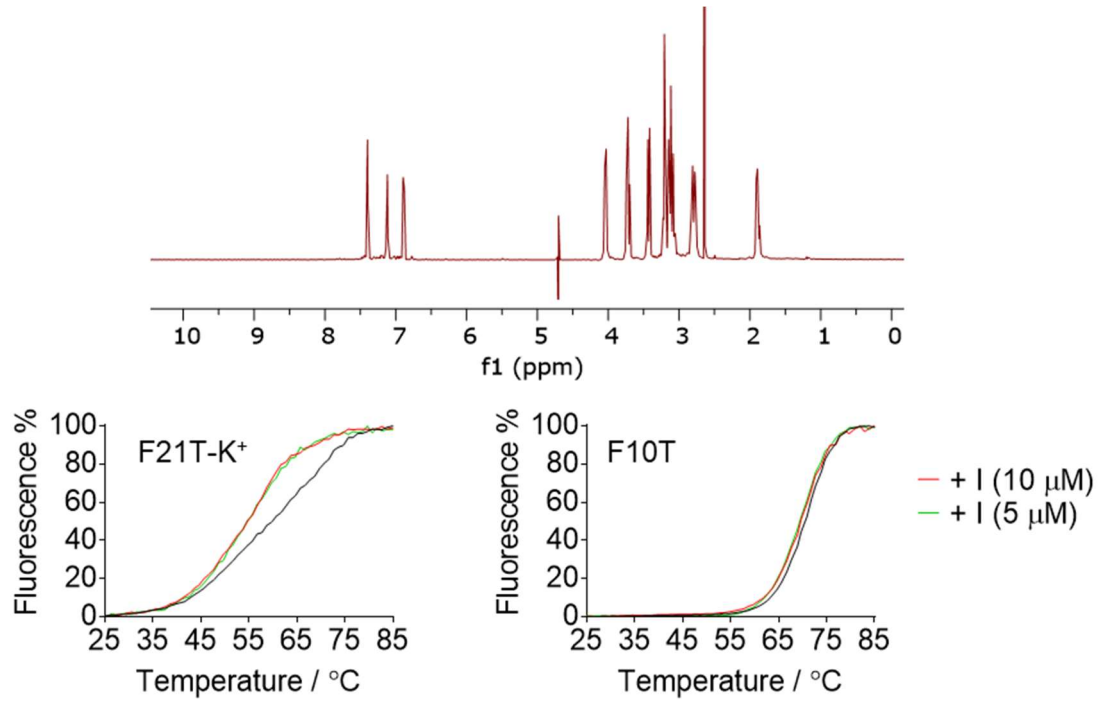


Figure A2.9: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **I**

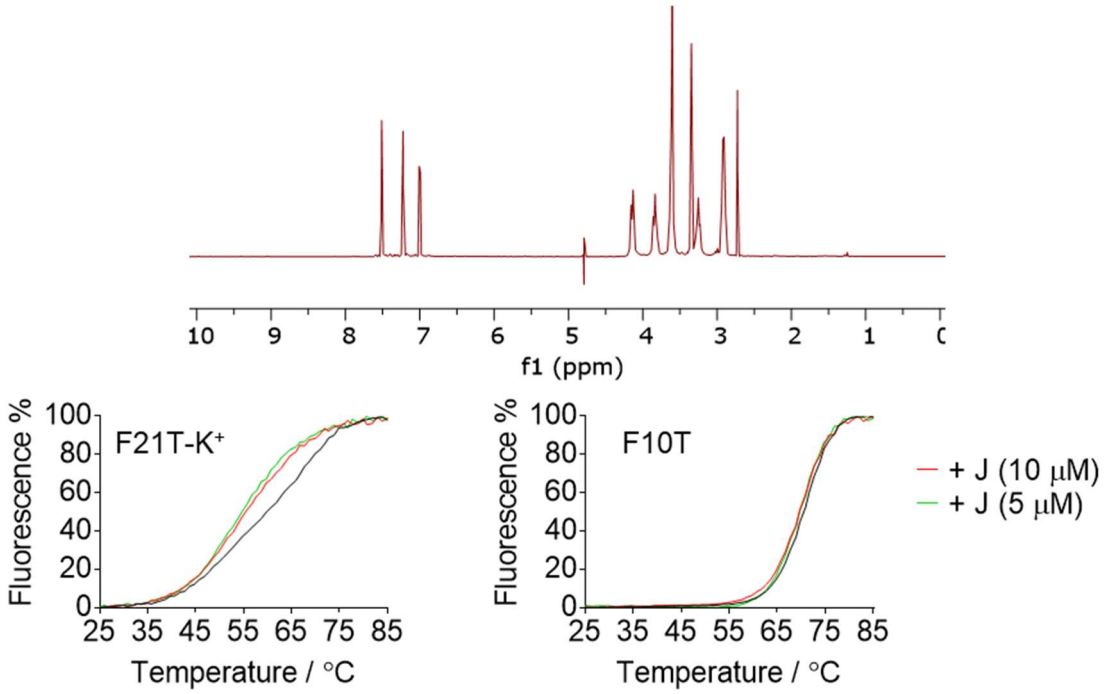


Figure A2.10: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **J**

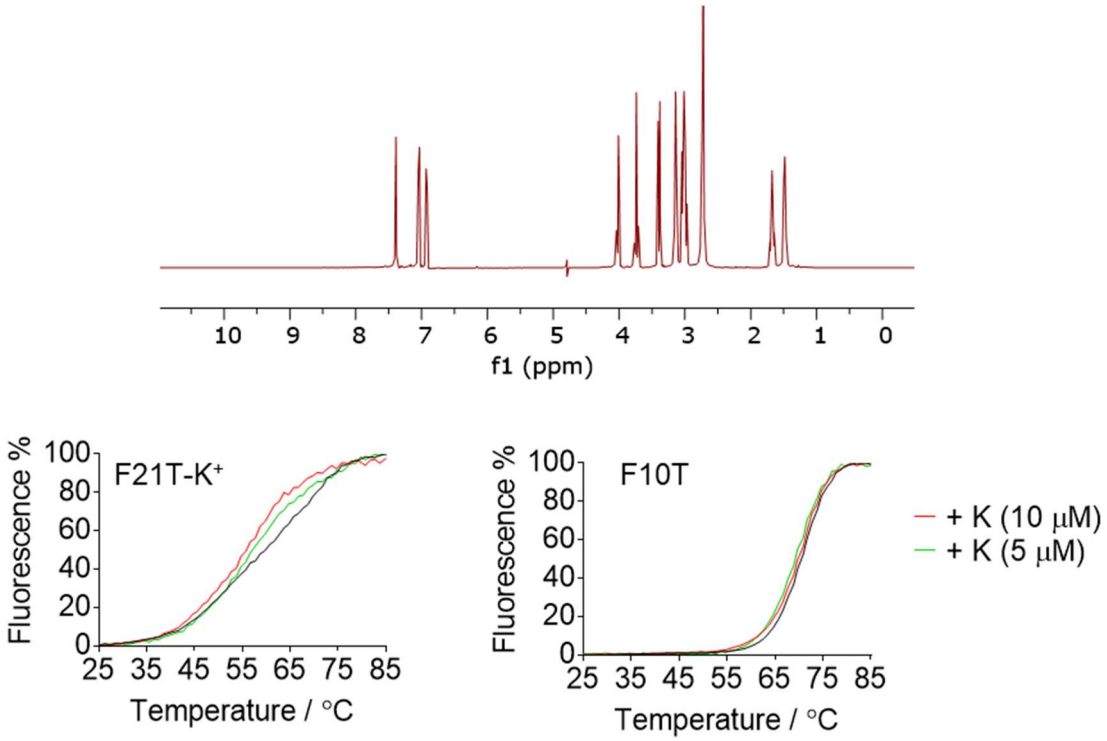


Figure A2.11: ^{1}H NMR spectrum (D_2O) and representative FRET curves for ligand **K**

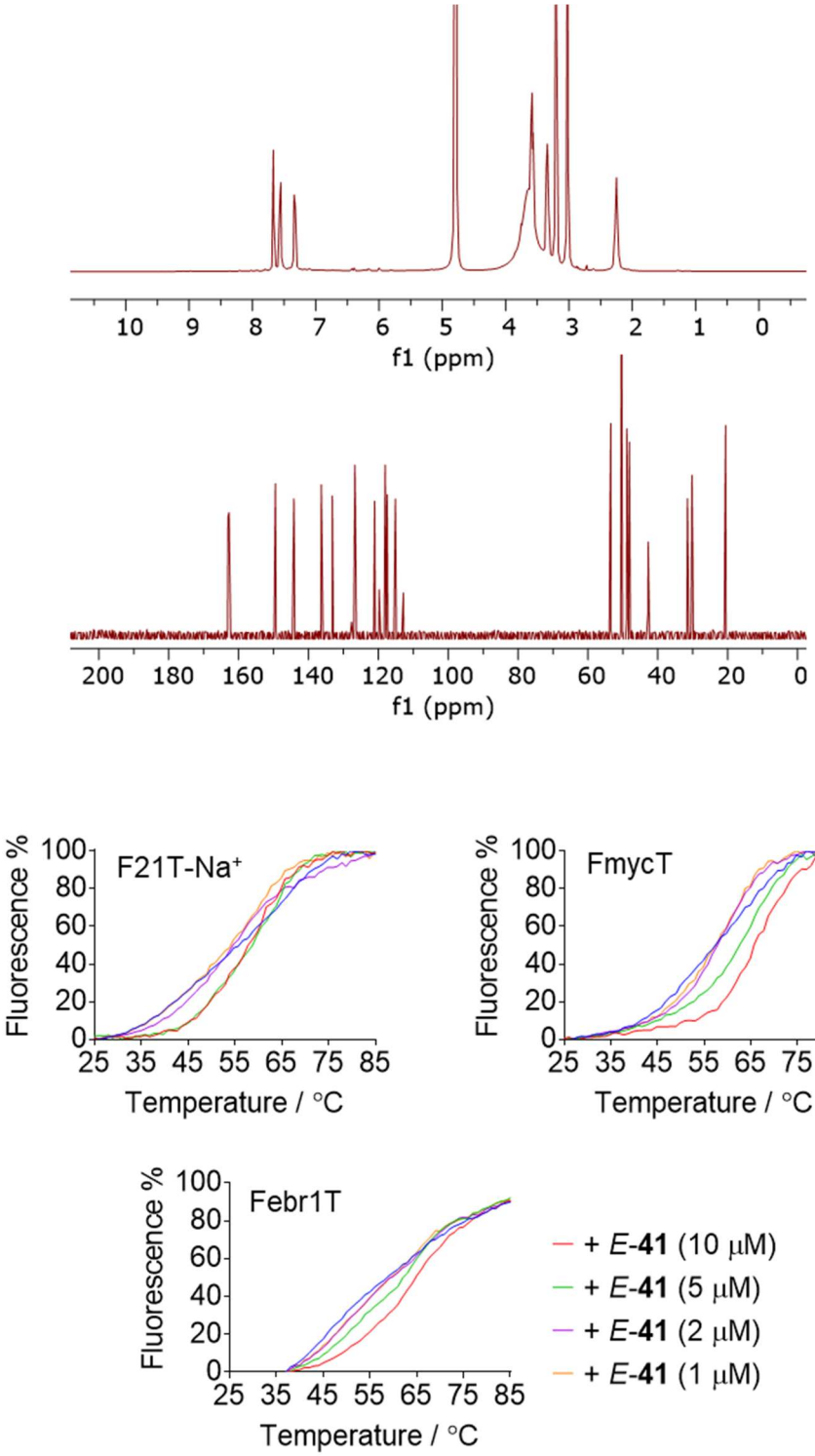


Figure A2.12: ^{1}H and ^{13}C NMR specta (D_2O) and representative FRET curves for ligand *E*-**41** (curves for F21T-K^+ and F10T are shown in Chapter 2).

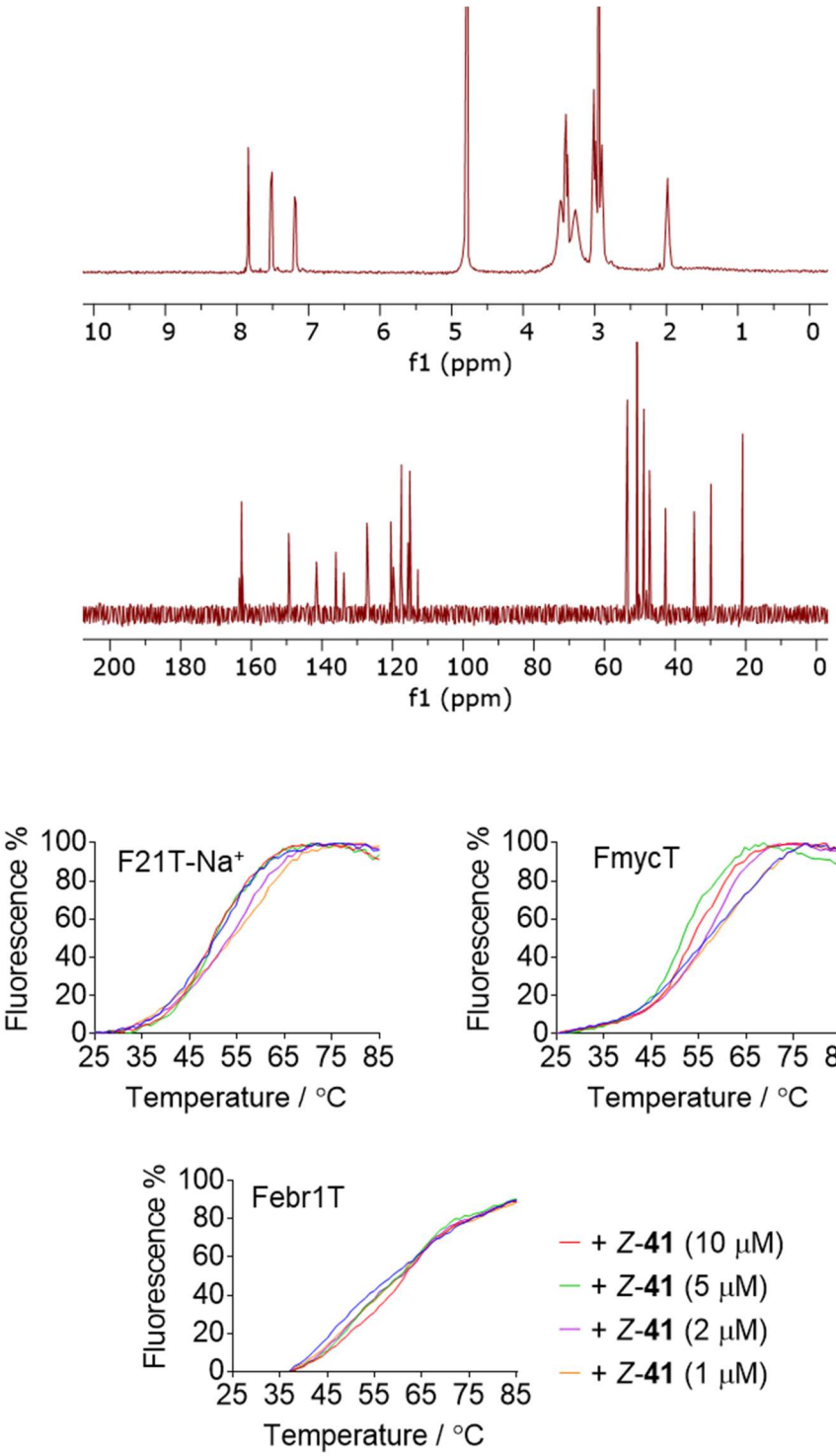


Figure A2.13: ^{1}H and ^{13}C NMR specta (D_2O) and representative FRET curves for ligand *Z*-**41** (curves for F21T-K^+ and F10T are shown in Chapter 2).

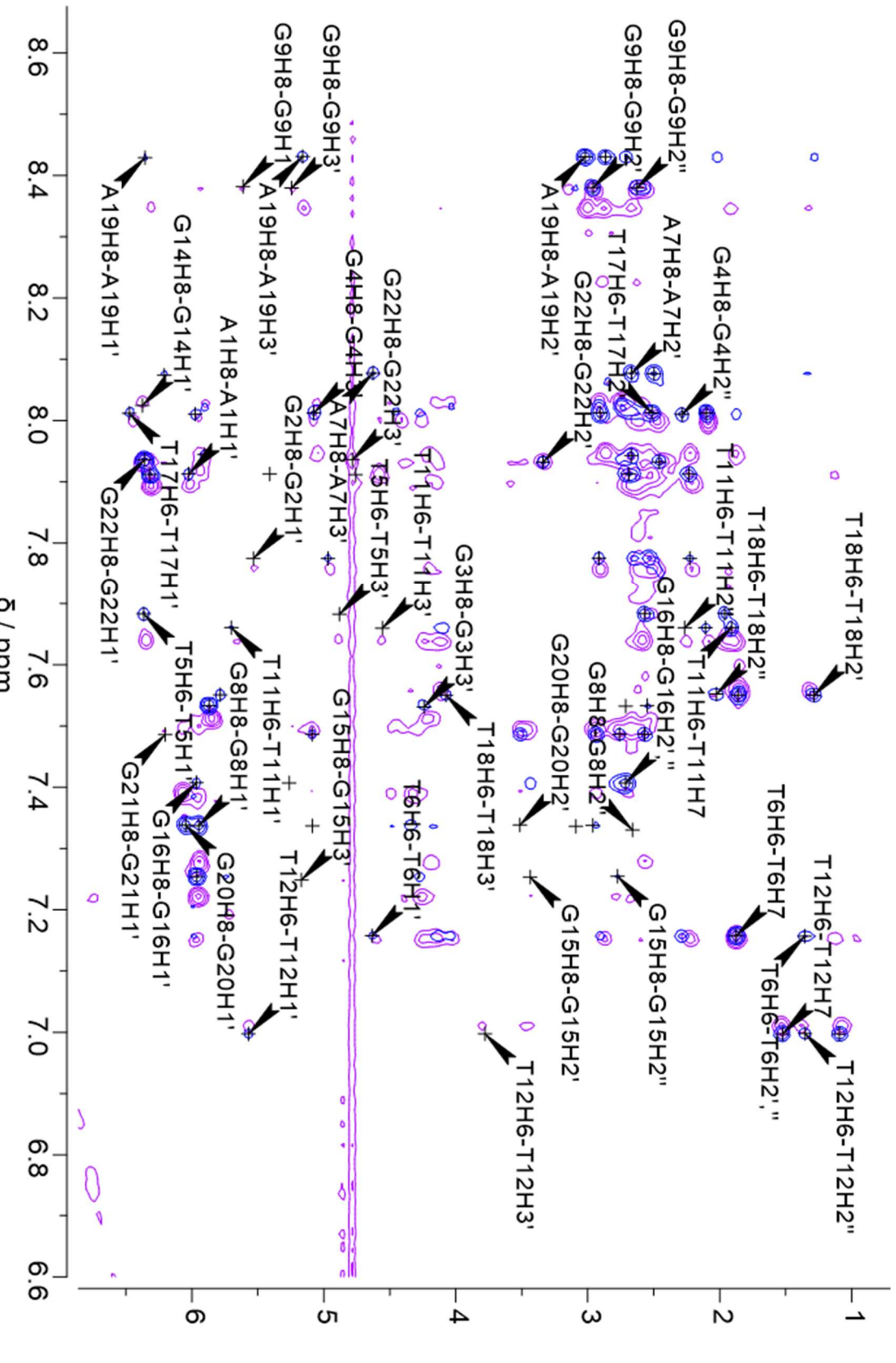


Figure A2.14: Overlaid NOESY spectra of telo22-Na^+ in the absence (blue) and presence (pink) of 2 equiv. ligand ***E*-41**

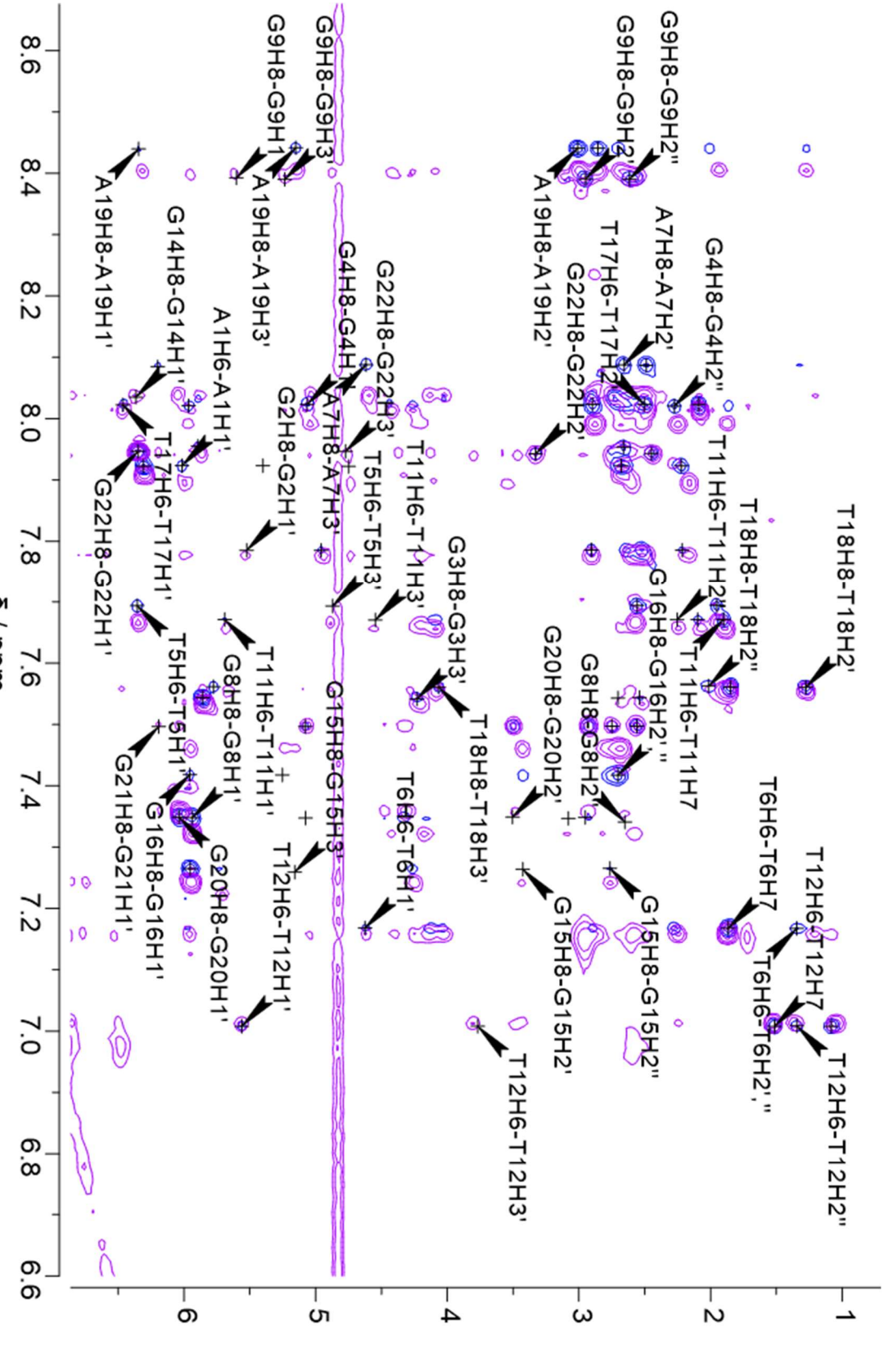


Figure A2.15: Overlaid NOESY spectra of telo22-Na^+ in the absence (blue) and presence (pink) of 2 equiv. ligand **Z-41**

A3 Supplementary data for Chapter 3

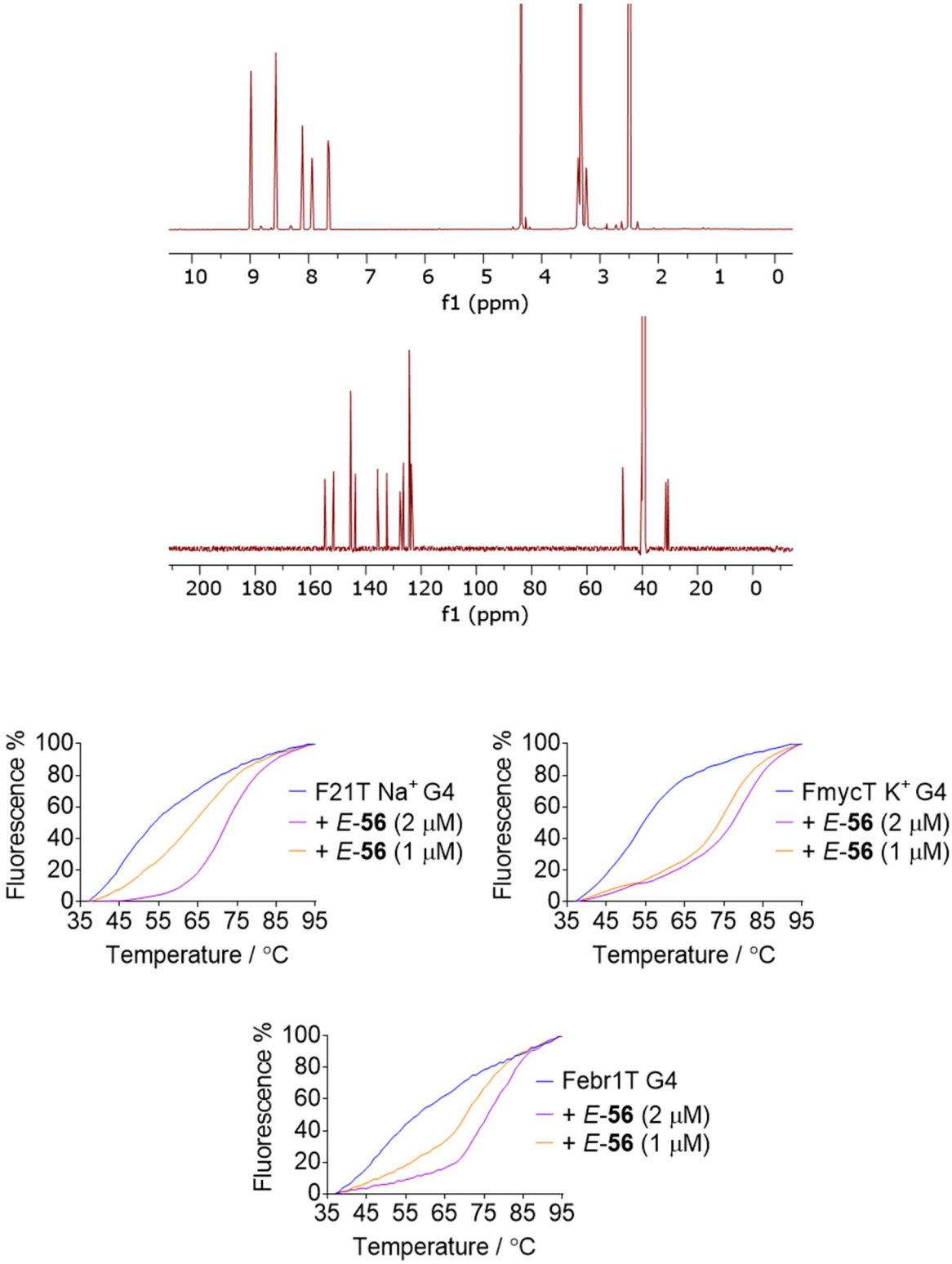


Figure A3.1: ^{1}H and ^{13}C NMR specta (DMSO-d_6) and representative FRET curves for ligand *E*-**56** (curves for F21T-K^+ and F10T are shown in Chapter 2).

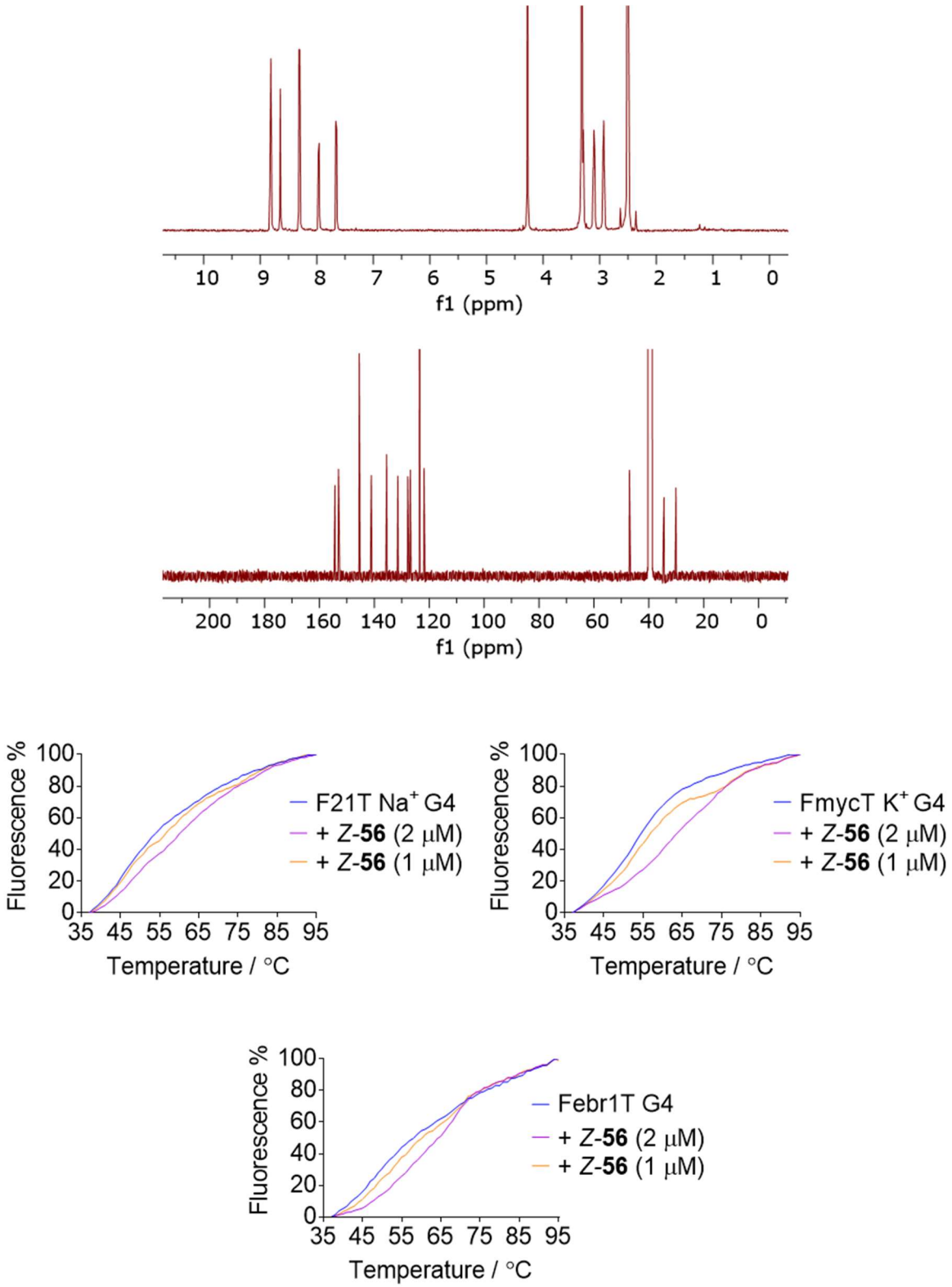


Figure A3.2: ^{1}H and ^{13}C NMR specta (DMSO-d_6) and representative FRET curves for ligand *Z*-**56** (curves for F21T-K^+ and F10T are shown in Chapter 2).

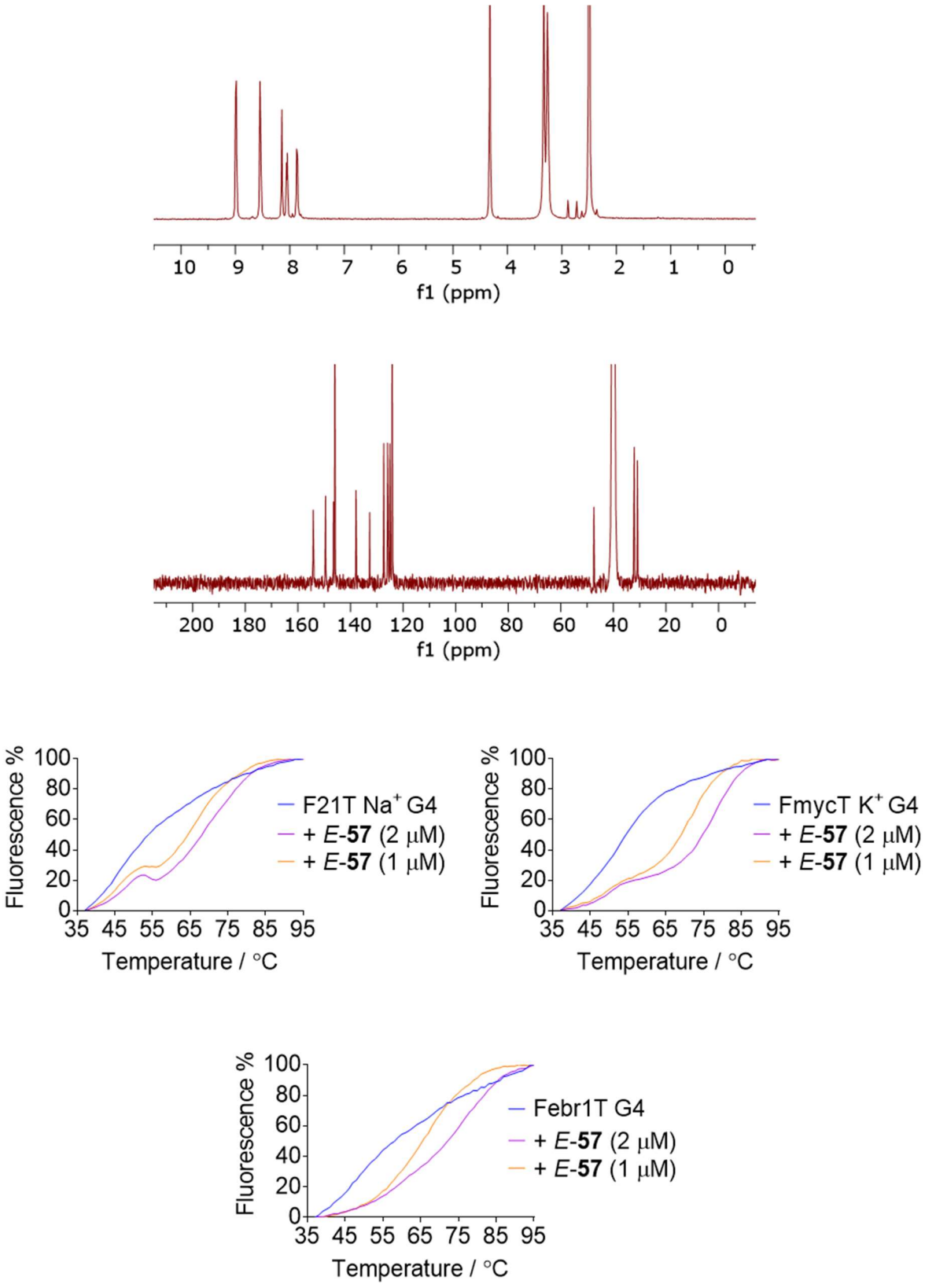


Figure A3.3: ^{1}H and ^{13}C NMR specta (DMSO-d_6) and representative FRET curves for ligand *E*-**57** (curves for F21T-K^+ and F10T are shown in Chapter 2).

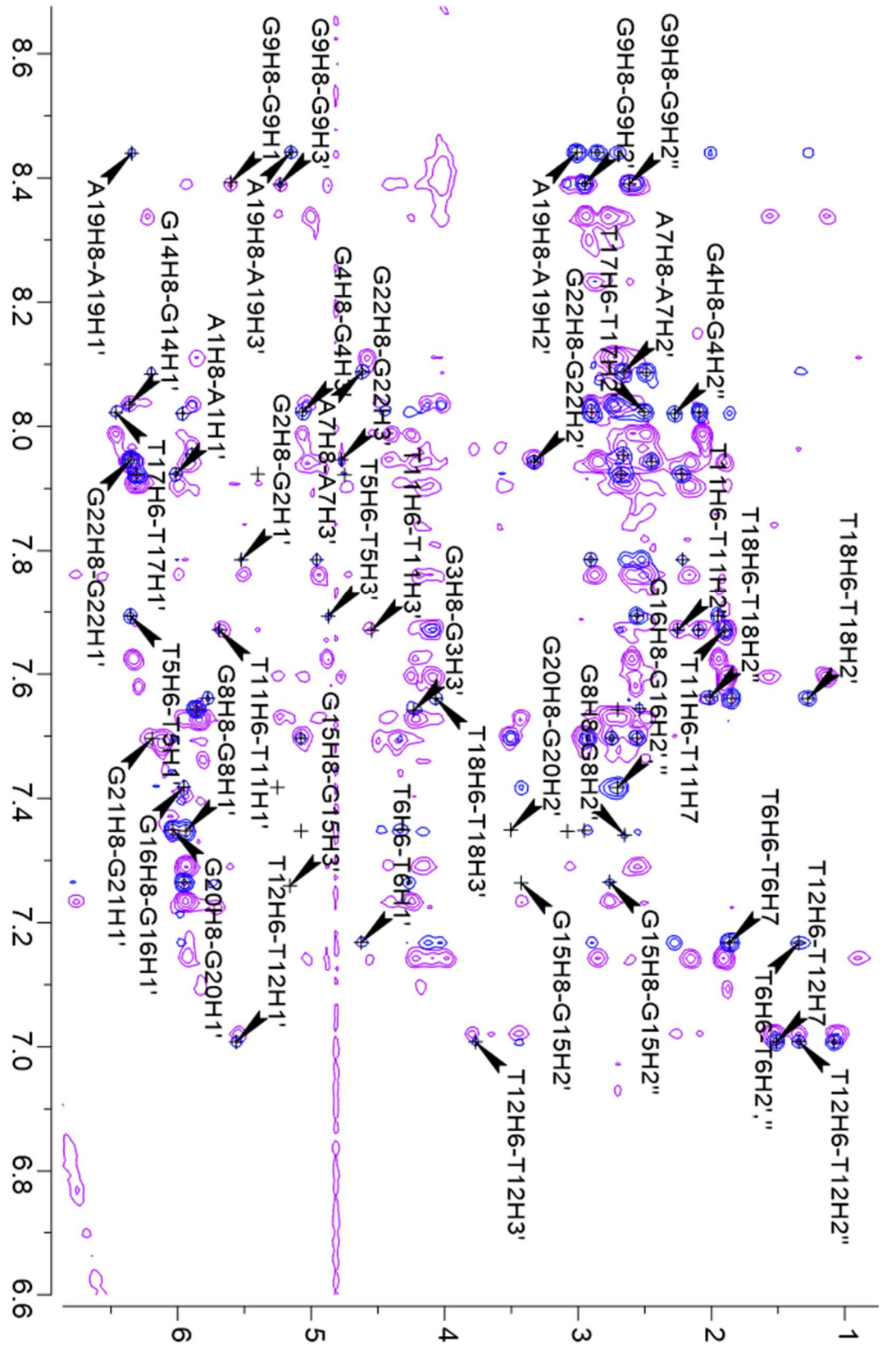


Figure A3.4: Overlaid NOESY spectra of telo22-Na^+ in the absence (blue) and presence (pink) of 2 equiv. ligand **Z-56**

A4 Supplementary data for Chapter 4

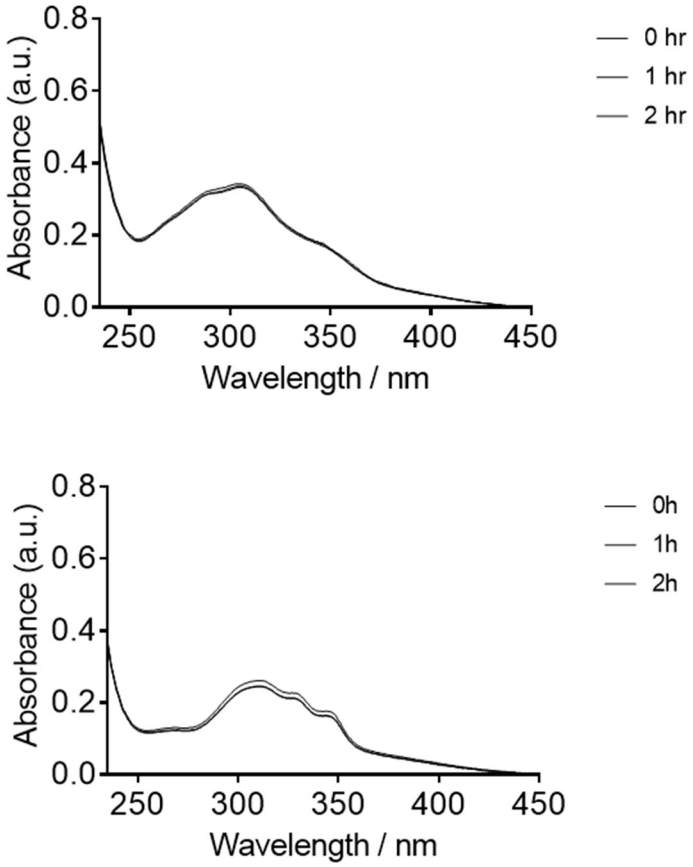


Figure A4.1: UV/visible spectra of (top) *Z*-**56** and (bottom) *E*-**56** after specified time points in the dark in 100 mM sodium phosphate buffer, pH = 7.4 showing stability of the ligand in absence of photoirradiation.

A5 Supplementary data for Chapter 5

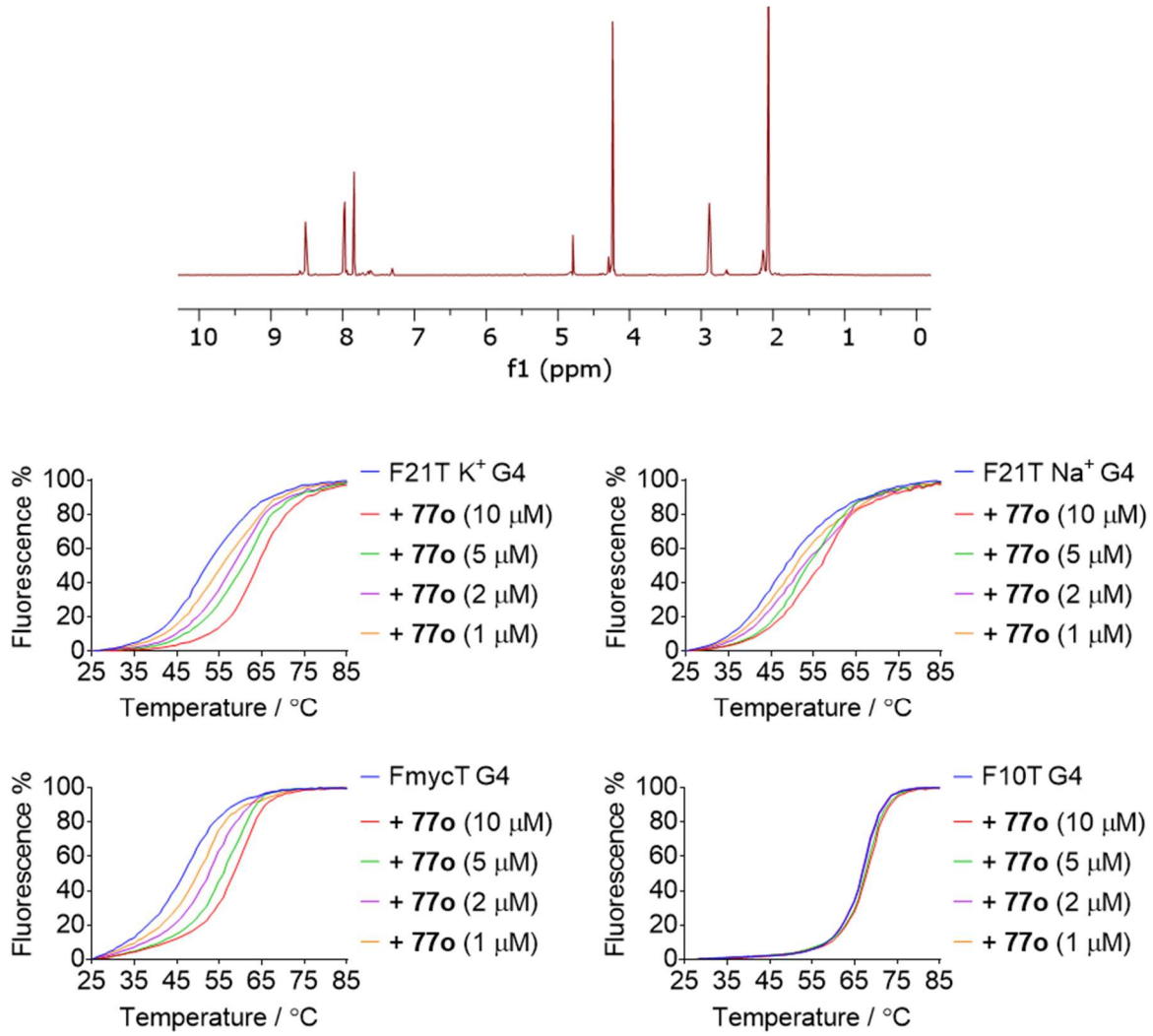


Figure A5.1: ^{1}H spectrum (D_2O) and representative FRET curves for ligand **77o** .

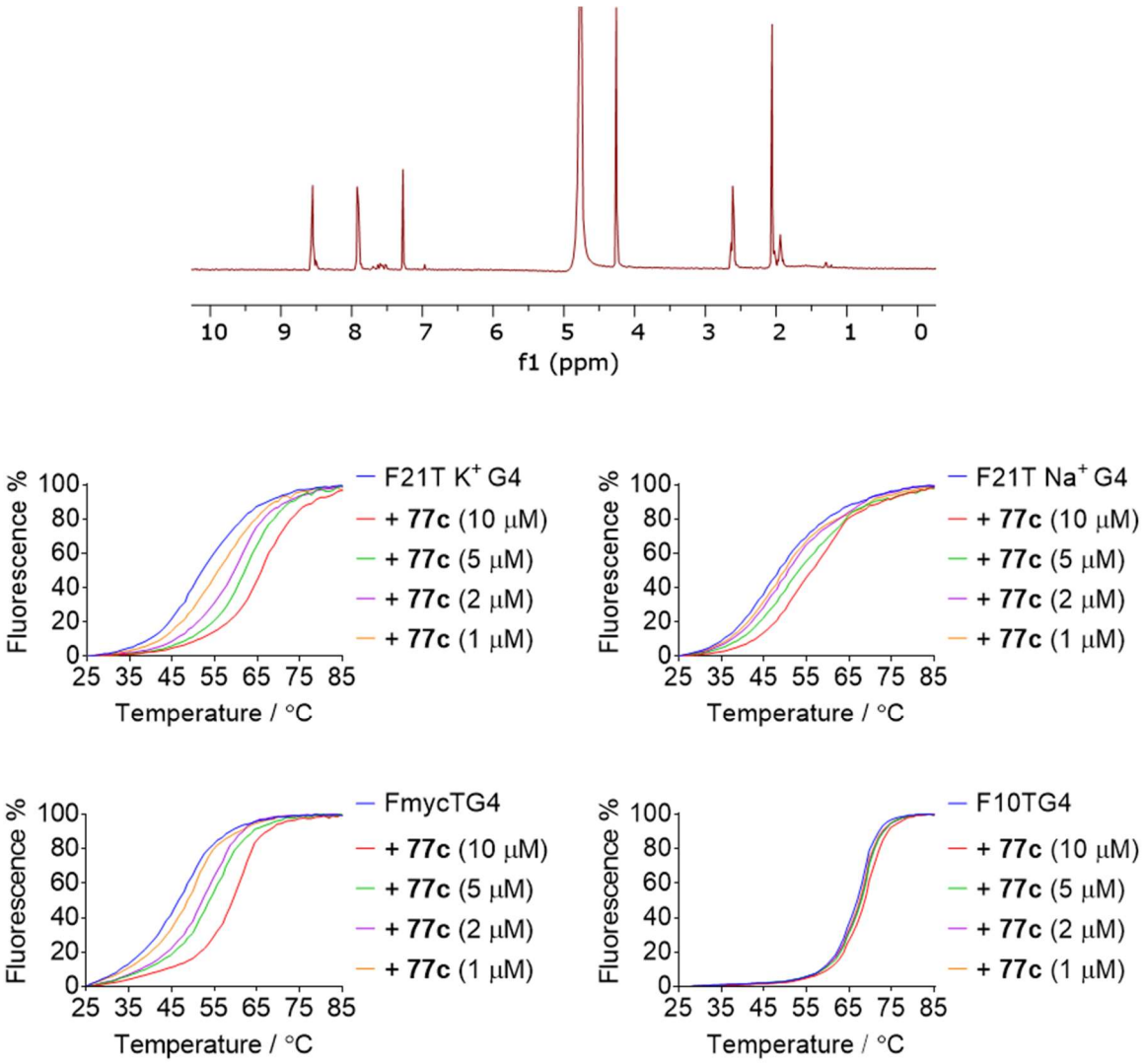


Figure A5.2: ^{1}H spectrum (D_2O) and representative FRET curves for ligand **77c** .

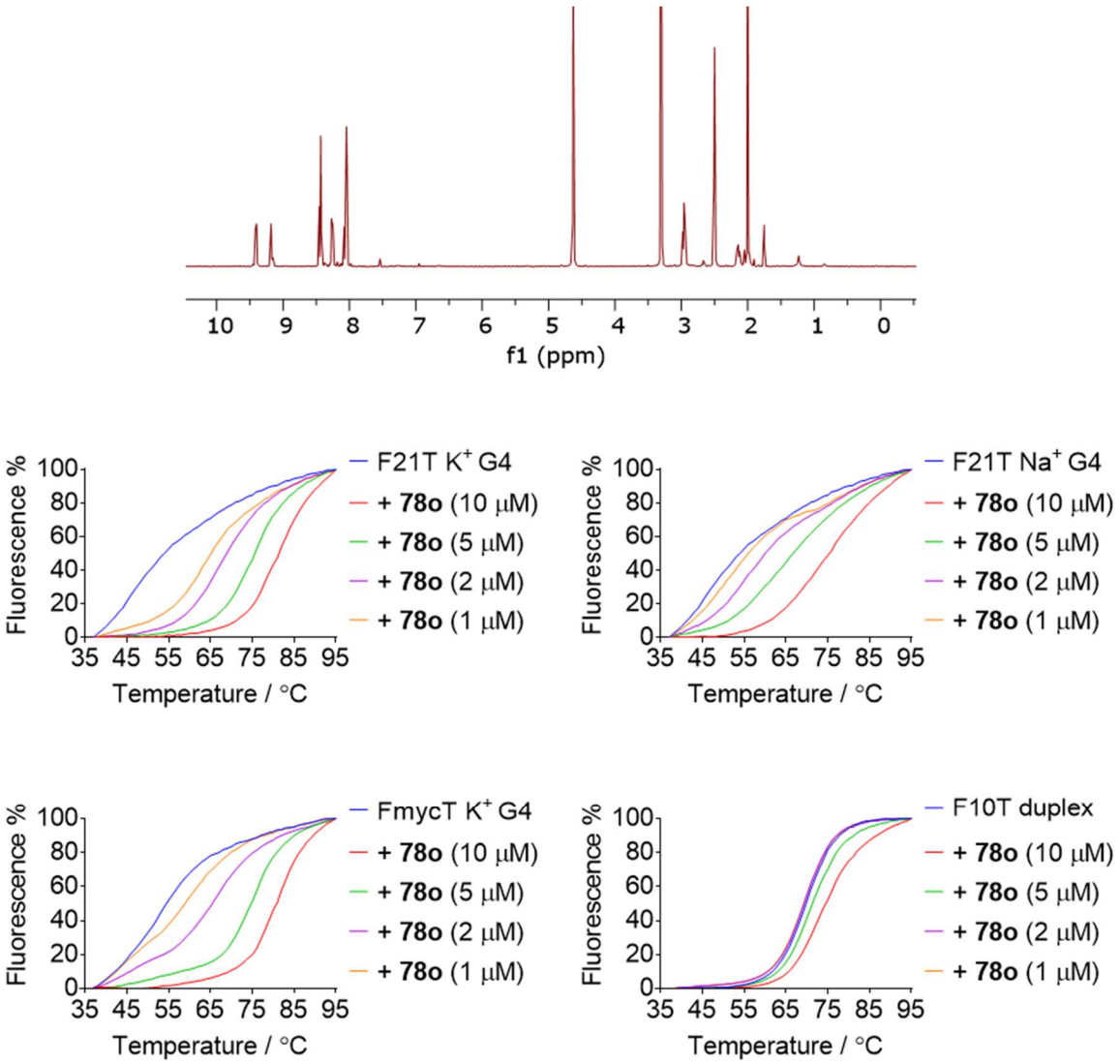


Figure A5.3: ^{1}H spectrum (DMSO-d_6) and representative FRET curves for ligand **78o** .

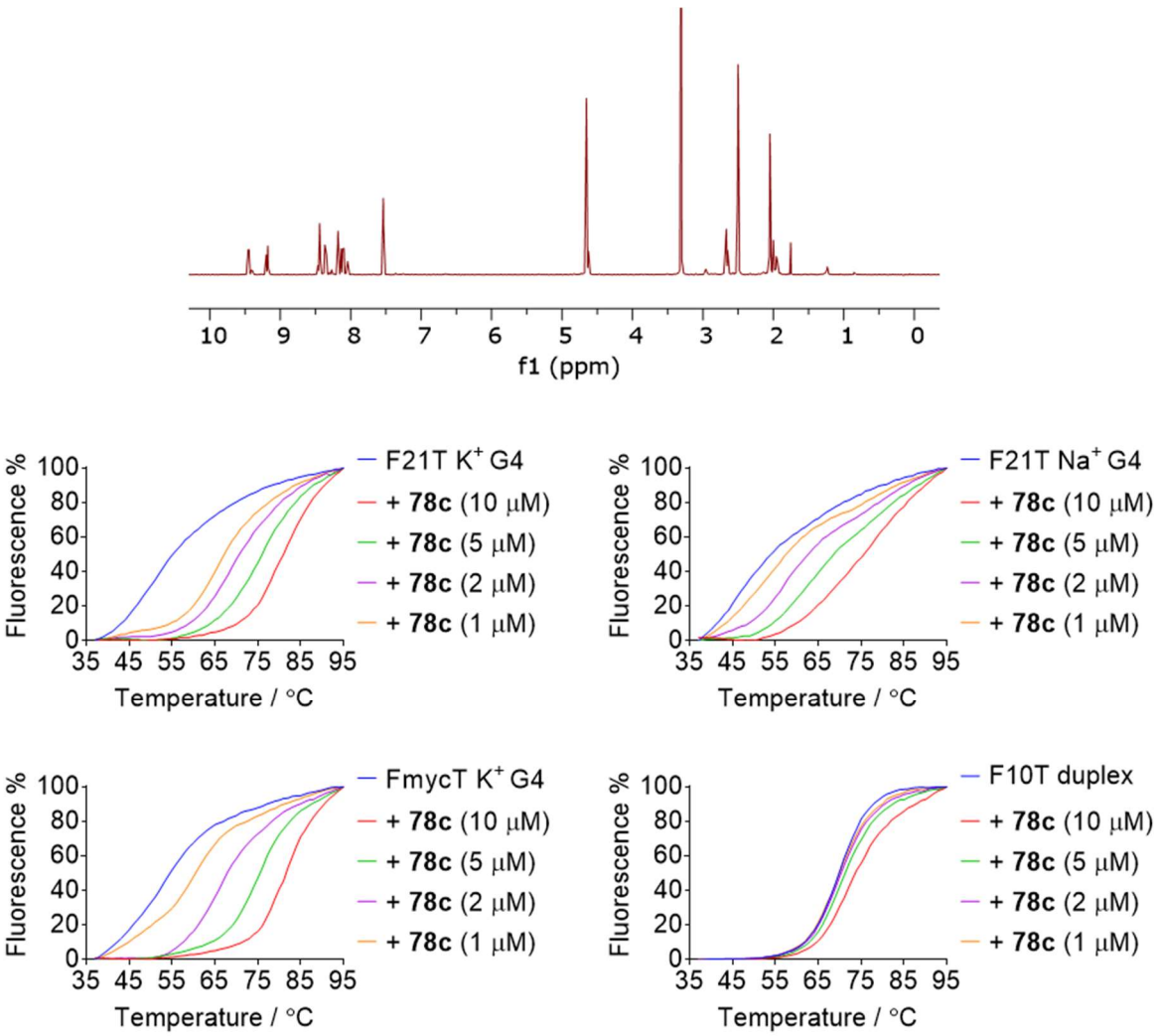


Figure A5.4: ^{1}H spectrum (DMSO-d_6) and representative FRET curves for ligand **78o** .

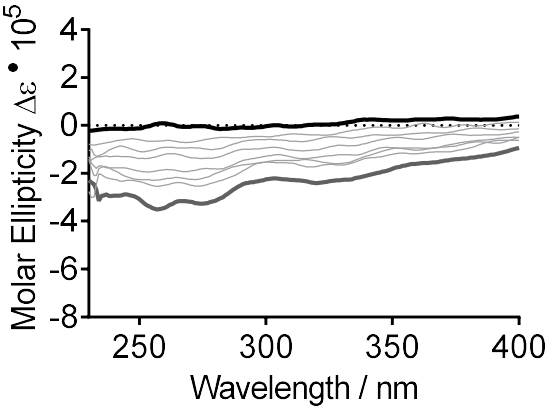


Figure A5.5: CD titration of ligand **78c** into 100 mM sodium phosphate buffer (100 mM, pH = 7.4) showing distortion of baseline proposed to arise from ligand aggregation. Black trace) buffer only; dark grey trace 30 μM ligand **78c**. Intermediate titration points are shown as light grey traces.

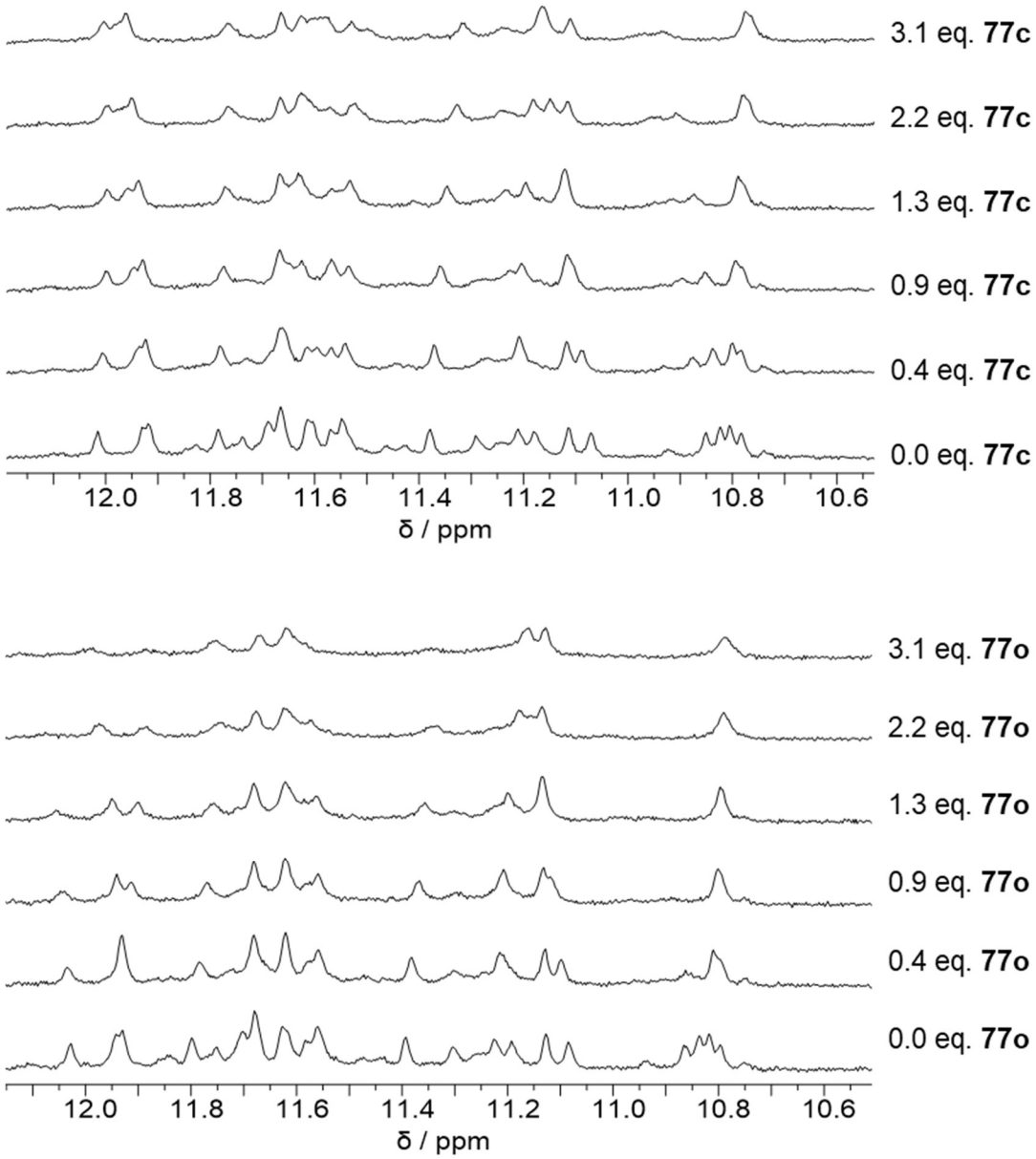
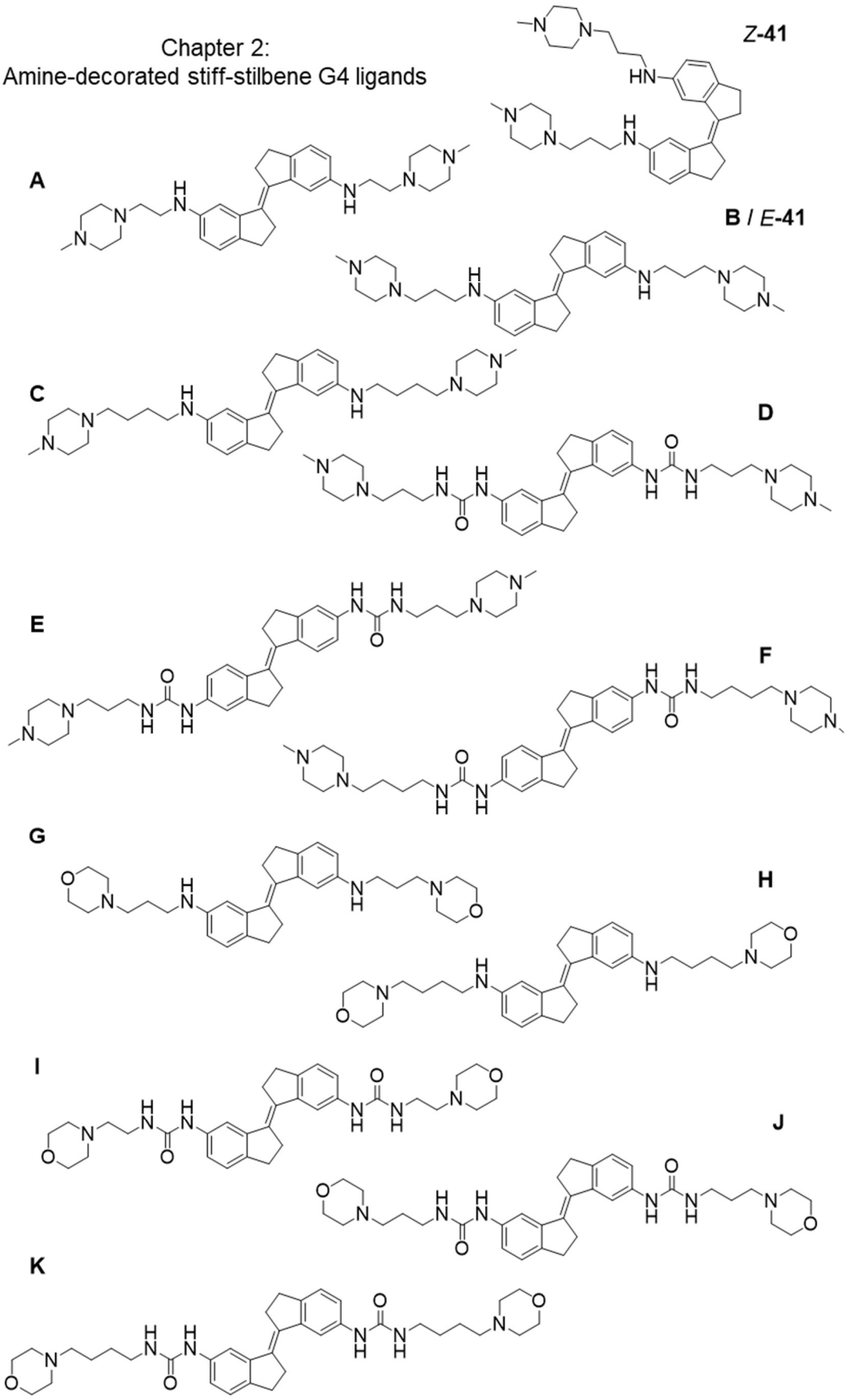
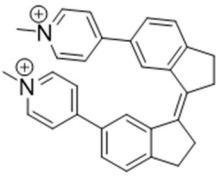


Figure A5.6: Stacked ^{1}H NMR spectra (imino region) of telo23-K^{+} G4 with increasing equivalents of (top) **77c** and (bottom) **77o**. DNA concentration: 185 µM.

Chapter 2:
Amine-decorated stiff-stilbene G4 ligands

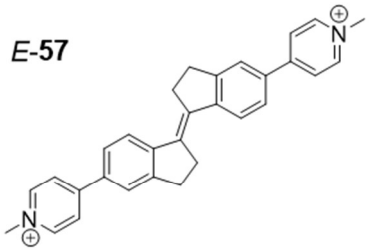


Chapter 3 & 4
N-methylpyridinium decorated
stiff-stilbene G4 ligands

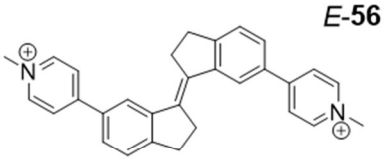


Z-**56**

E-**57**

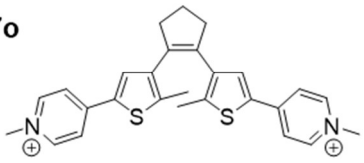


E-**56**

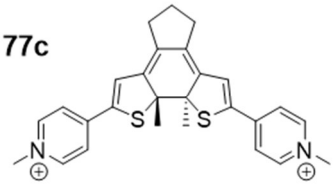


Chapter 5
Dithienylethene G4 ligands

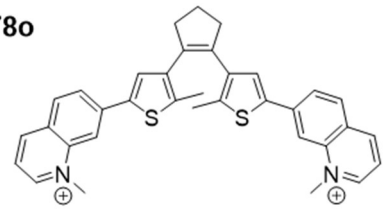
77o



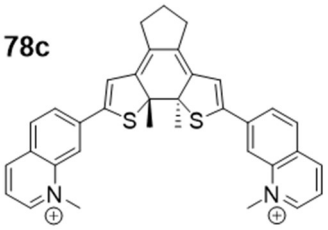
77c



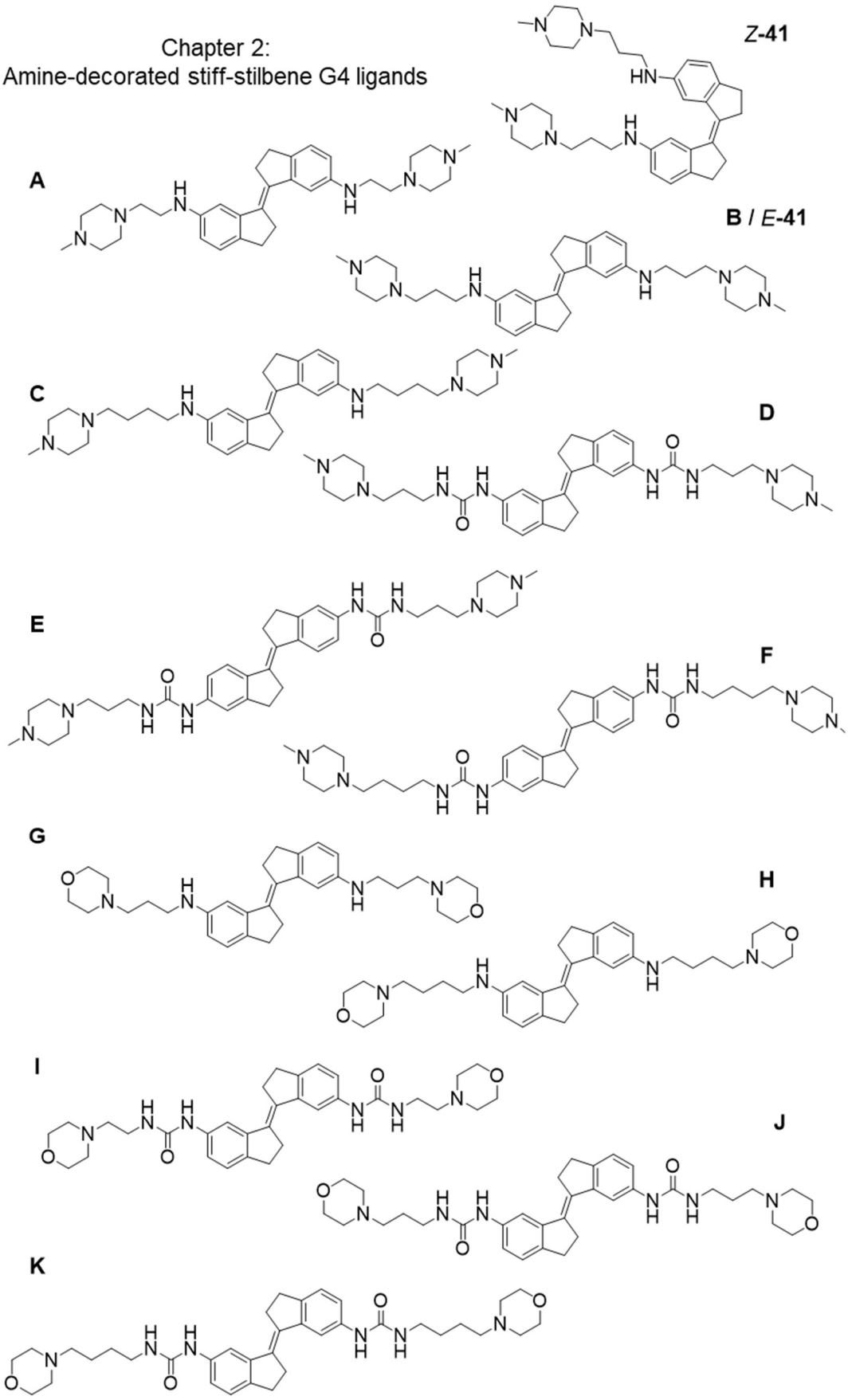
78o



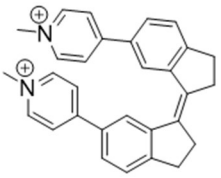
78c



Chapter 2:
Amine-decorated stiff-stilbene G4 ligands

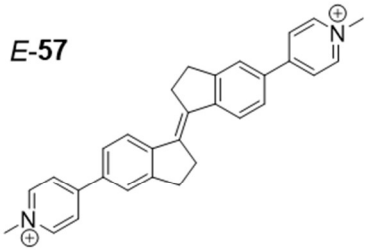


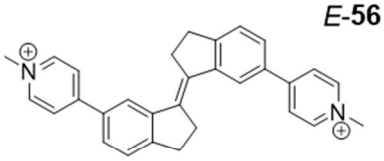
Chapter 3 & 4
N-methylpyridinium decorated
stiff-stilbene G4 ligands



***Z*-56**

***E*-57**

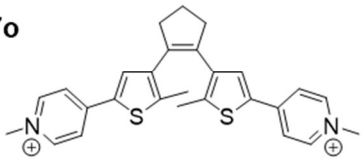




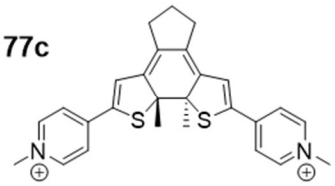
***E*-56**

Chapter 5
Dithienylethene G4 ligands

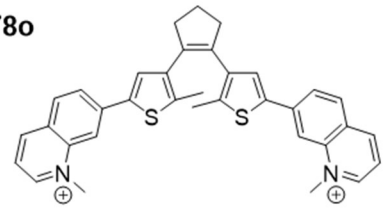
77o



77c



78o



78c